REMARKS

Claims 1, 2, 5-12, 14-26 and 37-59 are pending. No claims have been amended.

Claim Rejections – 35 USC § 103

The Examiner has rejected all claims under 35 U.S.C. § 103(a) as being unpatentable over Anderson et al. (U.S. Patent 5,922,591; "Anderson" hereinafter) in view of Waller et al. (Applied Environmental Microbiology, 2000, vol. 66, no. 9, pages 4115-4118; "Waller" hereinafter) as evidenced by Hassibi et al. (US Patent Publication No. 2004/0197845; "Hassibi" hereinafter).

Applicants' claims relate to a microfluidic device having an immunocapture chamber and a DNA analysis chamber integrated on the microfluidic device. A pneumatically actuated pump is also integrated on the device and configured to pump fluid from the immunocapture chamber into the DNA analysis chamber. The immunocapture chamber is "operable to capture a target provided to the immunocapture chamber through a microfluidic channel." Applicants' invention is able to provide effective mixing of antibody and antigen in a microfluidic device, which is required to achieve efficient immunocapture, as well as deliver the captured target to the DNA analysis chamber.

Anderson describes a miniaturized integrated nucleic acid diagnostic device, which has a number of reaction, storage or analytical chambers disposed within a single unit. Anderson does not teach or suggest an immunocapture chamber, or a chamber operable to capture an immunocapture target. Waller describes a macroscopic immunocapture process in a tube or other similar macroscopic (mL-sized) container.

Immunocapture processes performed at the macroscopic level in tubes cannot be extrapolated to microfluidic devices. Prior to Applicants' invention, the inability to transform macroscopic immunocapture processes (such as the Waller process) to micron-scale microfluidic devices was a barrier for the application of microfluidic systems to trace pathogen detection.

In rejecting the claims, the Examiner relies on Hassibi for proposition that "a microfluidic device comprising an immunocapture module was well known in the art at the time [the] invention was filed." (Office Action, page 11). Applicants respectfully traverse. Applicants' claims are supported by U.S. Provisional Patent Application No. 60/437,262 and so have an effective filing date of December 30, 2002. Applicants' effective filing date is prior to the July 24, 2003 filing date of Hassibi; accordingly Hassibi not prior art under 35 USC § 102(e). Applicants have reviewed the priority documents in Hassibi that have a filing date prior to

Applicants' effective filing date, namely U.S. Provisional Patent Application Nos. 60/407,412; 60/433,439; 60/435,924; and 60/435,934.[1] None of these priority documents disclose the subject matter on which the Examiner relies in the § 103 rejection.

At least because the references that are prior art under 35 USC § 102 do not teach or suggest an immunocapture chamber integrated on a microfluidic device, Applicants submit that the claims are patentable over the cited art.

Conclusion

If prosecution of this application can be assisted by telephone, the Examiner is requested to call Applicant's undersigned attorney at (510) 663-1100.

If any fees are due in connection with the filing of this amendment (including any fees due for an extension of time), such fees may be charged to Deposit Account No. 504480 (Order No. UCALP031).

Dated: December 16, 2008

Respectfully submitted,
Weaver Austin Villeneuve & Sampson LLP

/Denise S. Bergin/

Denise Bergin
Reg. No. 50,581

P.O. Box 70250
Oakland, CA 94612-0250

[1] Applicants are submitting copies of these references with this paper for the Examiner's convenience.

PATENT

Attorney Docket: 005852.P003Z

PROVISIONAL APPLICATION FOR UNITED STATES LETTERS PATENT

for

METHODS FOR EXPRESSION PROFILING AND DRUG DISCOVERY

by

Nader Pourmand

and

Arjang Hassibi

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 Wilshire Boulevard, Seventh Floor
Los Angeles, California 90025-1030

Telephone: (303) 740-1980
Facsimile: (303) 740-6962

EXPRESS MAIL CERTIFICATE OF MAILING

"Express Mail" Mailing No.: EV177134871US

I hereby certify that I am causing the above-referenced correspondence to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated below and that this paper or fee has been addressed to the Assistant Commissioner, Box Provisional Patent Application, Washington, D.C. 20231.

Date of Deposit:

Name of Person Mailing Correspondence:

Deborah Peloquin — Signature

August 30, 2002 — Date

Various methods of targeted drug discovery are known in the art. In some cases, those methods may comprise assays of gene expression levels in genetically engineered (knock-out) organisms and/or cells that have been exposed to one or more potential drugs or other therapeutic agents. A major problem with such present methods is that the dynamic range for determination of gene expression levels is limited to about two orders of magnitude.

The present invention discloses novel methods for expression profiling and drug discovery, using organisms and/or cells that have been engineered to express molecular barcodes. Methods of production of such barcoded cells are known in the art. The organisms and/or cells of use in the disclosed methods may also be genetically engineered to knock out at least one copy of one or more genes of interest.

In certain embodiments of the invention, a target cell may be prescreened to characterize gene expression levels for all genes expressed in the cell, tissue and/or organism. The target may or may not be first subjected to some treatment that may potentially change expression levels, such as hormonal treatment, exposure to an environmental condition or any other treatment that may affect gene expression levels. Gene expression may be measured by any technique known in the art. In certain embodiments, such techniques may include a Bioluminescence Regenerative Cycle (BRC) (*e.g.*, U.S. Patent Application Serial No. 10/186,455, filed June 28, 2002, the entire text of which is incorporated herein by reference.)

In various embodiments of the invention, the expressed nucleic acid products may be divided into groups, for example of high level expression, intermediate level expression and/or low level expression. Probes for such expressed nucleic acid products may be placed on separate chips for use in drug discovery assays utilizing gene expression. For example, a candidate drug may be selected on the basis of its ability to alter (increase or decrease) the expression level(s) of one or more selected genes. The methods disclosed herein are advantageous in overcoming the dynamic range limitations for nucleic acid chips used to assay gene expression levels. Additional details of the claimed methods are disclosed in the attached pages (the entire text of which is incorporated herein by reference).

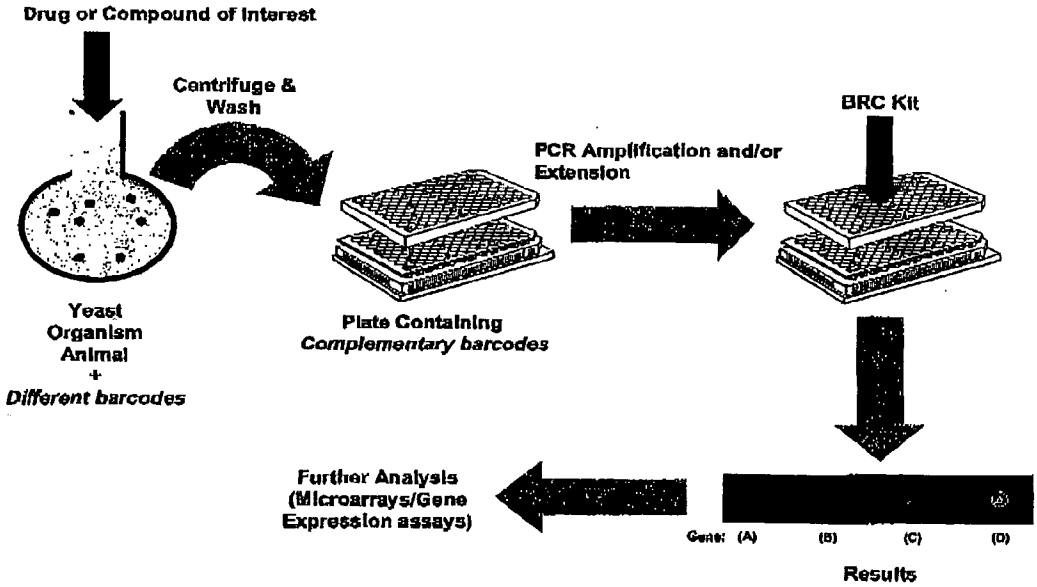
The skilled artisan will realize that various methods, compositions and apparatus that are well known in the art have not been recited herein for simplicity.

Authors: Nader Pourmand/Arjang Hassibi

Assay/Process/Method Patent: BRC application for Drug Discovery/ Mechanism of Action determination of known/unknown compounds

What we are describing in this provisional is a method to elucidate the mechanism of action of drugs or compounds by screening the gene expression patterns of all the genes in an organism to determine which gene or pathway was affected.

The process starts with drug/compound (Molecule) of interest. The Molecule of interest is incubated with organisms, each with unique barcode (yeast, mice cells, human cells-Cells). The Cells are centrifuged and washed. Cells with barcodes can be placed in a Microtiter plate (single or multiplexed). DNA/RNA is extracted. Each cell's DNA/RNA is placed individually or in multiplex fashion into Microtiter plate containing the complementary barcode. Depending on the number of copies of DNA/RNA, PCR can be performed or not. After, BRC reagents are added to each well in the microplate and the microplate is scanned with any light sensitive and/or light sensitive system (CCD camera, luminometer, photodiode, etc.). Microtiter-wells that are exposed will generate light. Those barcodes that are present in the well can be further evaluated using microarrays or other methods to measure the expression levels of target genes.



BRC application for drug discovery – Mechanism of action for compounds

Authors: Nader Pourmand/Arjang Hassibi

Method/Assay/Process Patent: Bioluminescence Super-Regenerative Cycle (BSRC)

In this patent we describe a method to increase light intensity generated by the BRC method (modified method);

In this method PPi is converted to ATP by the Enzyme, ATP Sulfurylase, and co-enzyme APS. ATP molecules are substrate for luciferase and combined with co-enzyme Luciferin generate light and PPi molecules. In addition, we have added to the reaction the enzymatic complex of Adenylate Kinase (AK) in the presence of AMP Substrate and pyruvate kinase (PK) in the presence of phosphoenolypyrvate (PEP). The two later enzymes (PK and AK) can potentially create two ATP molecules from a single ATP. This process would exponentially increase the concentration of ATP molecules in the reaction buffer. Bioluminescence light activity of luciferase is proportional to the ATP concentration. Thus the amount of light that we see from the assay, as a function of time, has an exponential growth and the rate of the growth depends on the initial PPi concentration and/or ATP. Applications include Gene Expression, Pathogen Profiling, SNP detection, biological warfare, immunoassay and cell based screenings.

SBRC

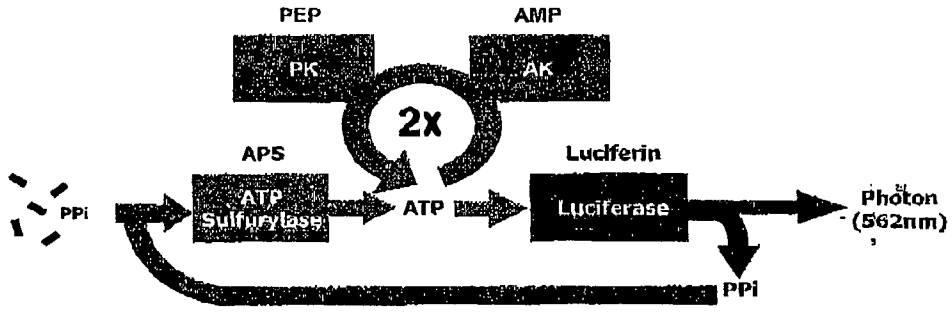
BRC

$$PPi + APS \xrightarrow{ATP-Sulfurylase} ATP + SO_4^{2-}$$

$$ATP + Luciferin + O_2 \xrightarrow{Luciferase} Oxyluciferin + AMP + CO_2 + h\nu + PPi$$

$$\boxed{ATP} + AMP \xrightarrow{AK} 2ADP$$

$$2ADP + 2PEP \xrightarrow{PK} 2Pyr + \boxed{2ATP}$$



Bioluminescence Super-Regenerative Cycle Concept

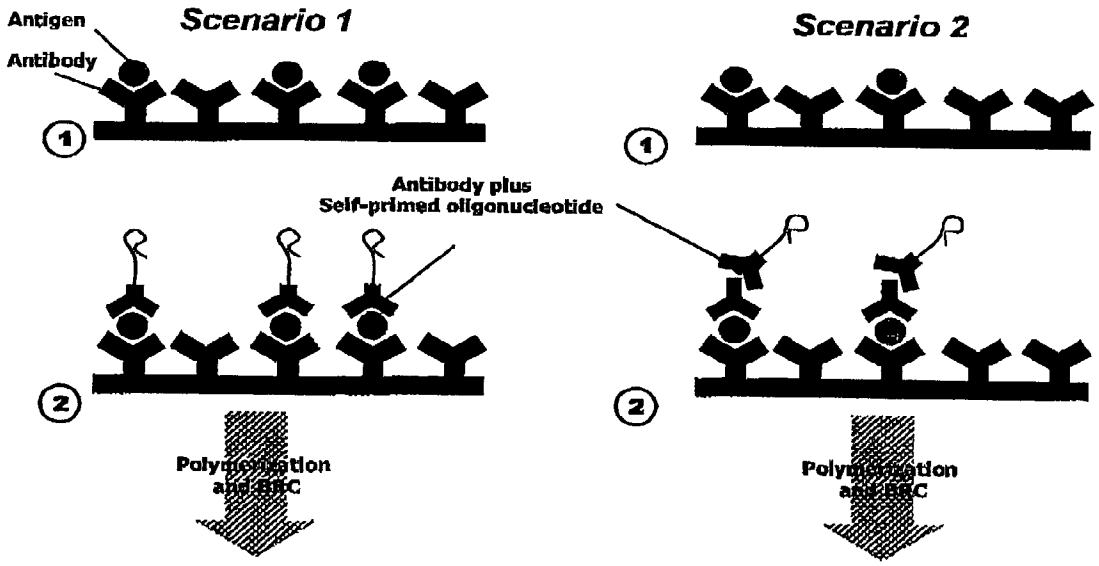
Authors: Nader Pourmand/Arjang Hassibi

Method/Process/Assay Patent: Immunoassays using BRC/BSRC method in solution and solid phase

In this patent we describe two methods for conjugating Antibody with either self-primed oligonucleotide or a luciferase enzyme and using it in immunoassay applications in solutions or solid phase

Scenario I: self-primed oligonucleotide linked to Antibody

Primary antibody is used to capture an antigen. This primary antibody assay can be performed in solution or localized on a solid surface. Once the antibody has bound to the antigen, the antibody attached to the antigen is washed to get rid of all remaining non-bound molecules. A second antibody is introduced with a self-primed oligonucleotide attached. The second antibody will detect the existing antibody-antigen complex and attach itself (this is the same as a standard ELISA or Antibody Sandwich Assay). The self-primed oligonucleotide serves as template for BRC or BSRC assay.

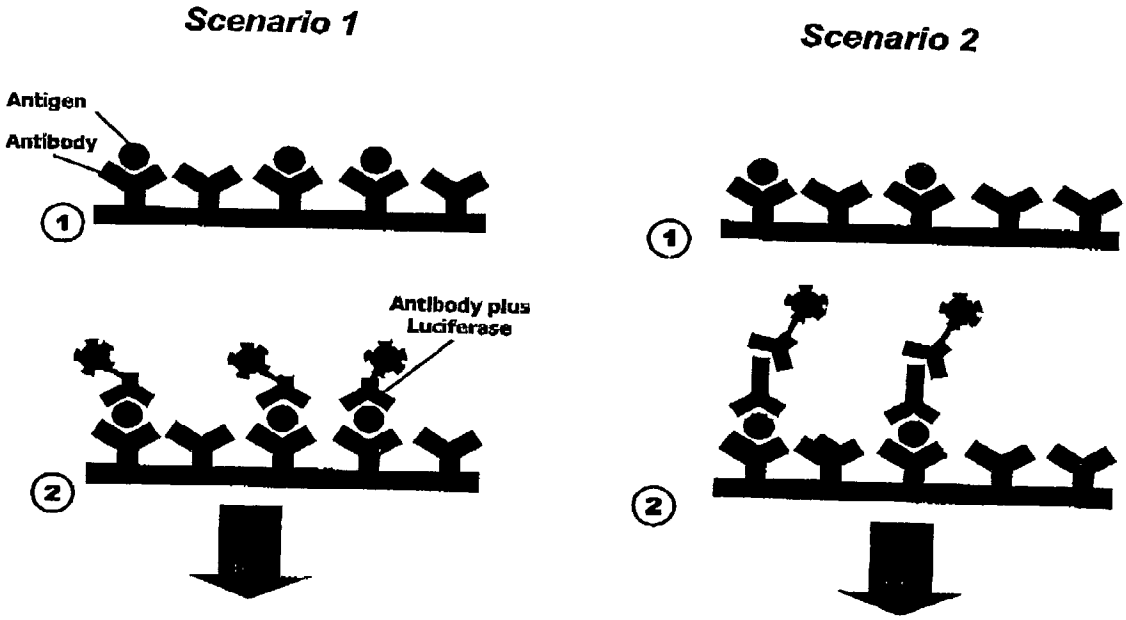


BRC Immunoassay with self-primed oligonucleotide attached to antibody

Method/Process/Assay Patent: Immunoassays using BRC/BSRC method in solution and solid phase (Continued)

Scenario II: Luciferase linked to Antibody

Primary antibody is used to capture an antigen. This primary antibody assay can be performed in solution or localized on a solid surface and multiplexed. Once the antibody has bound to the antigen, the antibody attached to the antigen is washed to get rid of all remaining non-bound molecules. A second antibody is introduced that attaches to a different site on the antigen if the antibody/Antigen is on a solid surface or multiplexed in an array format. A third antibody is introduced that is linked to Luciferase (different luciferase can potentially generate different colors). The Antibody linked to luciferase will bind to the second Antibody and serve as a template for BRC assay. BRC reagents minus luciferase are added to the solution generating light that can be different colors based on the fact that different luciferases can generate different light wavelengths when activated. Alternatively in solution there is no need for the second antibody. The antibody linked to luciferase will bind to the primary antibody attached to the antigen and serve as a template for the BRC assay (as in ELISA based assays).



BRC Immunoassay with Luciferase attached to antibody

PATENT

Attorney Docket: 005852.P004Z

APPLICATION FOR UNITED STATES LETTERS PATENT

for

BIO-MOLECULAR SEQUENCE RECOVERY USING STATISTICAL SIGNAL PROCESSING

by

Babak Hassibi

Arjang Hassibi

and

Nader Pourmand

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 Wilshire Boulevard, Seventh Floor
Los Angeles, California 90025-1030

Telephone: (303) 740-1980
Facsimile: (303) 740-6962

EXPRESS MAIL CERTIFICATE OF MAILING
"Express Mail" Mailing No.: EV 118157899 US

I hereby certify that I am causing the above-referenced correspondence to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated below and that this paper or fee has been addressed to the Assistant Commissioner, Box Provisional Patent Application, Washington, D.C. 20231.

Date of Deposit: October 29, 2002

Name of Person Mailing Correspondence: Lori E. True

Lori E. True
Signature

October 29, 2002
Date

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to the field of bio-molecule characterization based on statistical signal processing. More particularly, the present invention concerns novel approaches to identification, characterization and/or sequencing of biomolecules, such as nucleic acids or proteins, using a combination of the law of large numbers and Wiener filtering theory for statistical signal processing.

Description of Related Art

[0002] Existing methods for nucleic acid sequencing, based on detection of labeled nucleic acids that have been separated by size, are limited by the length of the nucleic acid that can be sequenced. Typically, only 500 to 1,000 bases of nucleic acid sequence can be determined at one time. This is much shorter than the length of the functional unit of DNA, referred to as a gene, which can be tens or even hundreds of thousands of bases in length. Using current methods, determination of a complete gene sequence requires that many copies of the gene be produced, cut into overlapping fragments and sequenced, after which the overlapping DNA sequences may be assembled into the complete gene. This process is laborious, expensive, inefficient and time-consuming.

[0003] More recent methods of nucleic acid sequencing, involving hybridization to oligonucleotide arrays of known sequences at specific locations on a chip, may be used to infer short nucleic acid sequences or to detect the presence of a specific nucleic acid in a sample. However, they are not suited for identifying long nucleic acid sequences.

[0004] Pyrophosphate based detection systems have been used for DNA sequencing (*e.g.*, Nyren and Lundin, *Anal. Biochem.* 151:504-509, 1985; U.S. Patent Nos. 4,971,903; 6,210,891; 6,258,568; 6,274,320, each incorporated herein by reference). The method uses a coupled reaction wherein pyrophosphate is generated by an enzyme-catalyzed process, such as nucleic acid polymerization. The pyrophosphate is used to produce ATP, in an ATP sulfurylase catalyzed reaction with adenosine 5'-phosphosulphate (APS). The ATP in turn is used for the production of light in a luciferin-luciferase coupled reaction. Although an improvement on earlier methods of nucleic acid sequencing, this technique still does not address the problem of

data analysis for complex signals generated during the sequencing of large numbers of template nucleic acid molecules. Improved methods applicable to data analysis during biomolecule sequencing are needed.

SUMMARY OF THE INVENTION

[0005] The present invention provides novel methods for the recovery of an information-bearing *signature* signal from the aggregate signal obtained from the random superposition of a very large number of these signatures. This problem arises in many different applications and the skilled artisan will realize that the disclosed methods have broad applicability in a variety of fields. Particular embodiments of the invention discussed below deal with the applications in the area of genomics, especially in the area of rapid real-time DNA sequencing.

[0006] Although from a mathematical point of view the data analysis problem can be categorized in the general area of inverse problems, what distinguishes the problem in genomic applications is the sheer number of signature signals that form the aggregate signal. For example, a similar problem arises in wireless communications where the information-bearing electromagnetic signal arrives at the receiver with random delays, due to different reflection and scattering paths from the transmitter to the receiver. The receiver must recover the information-bearing signal from this aggregate signal. Whereas in wireless communications the number of superposed signals may range from the tens to the several hundreds, in genomic applications the number of superposed signals is equal to the number of DNA molecules present in a given sample. This number may range from several hundred thousand to trillions or more. Therefore the techniques employed in wireless communications, such as the estimation of the different delays, etc., are not feasible for genomic applications. Resolution of the data analysis problem, using the methods disclosed herein, allows the inexpensive real-time sequencing of biomolecules, such as nucleic acids or proteins.

[0007] In certain preferred embodiments, nucleic acid sequencing signals are generated by a novel method referred to as "bioluminescence regenerative cycle" or BRC. The method may comprise obtaining at least one sample containing one or more template nucleic acids to be sequenced, generating pyrophosphate by replication of the template nucleic acid, producing light by a bioluminescence regenerative cycle, detecting a complex signal from the BRC process, deconvoluting the signal by statistical signal processing and determining the sequence of the

template nucleic acid. In particular embodiments, the template nucleic acid may be replicated by polymerase chain reaction amplification, although in alternative embodiments any process or reaction that results in the production of pyrophosphate may be coupled to BRC analysis.

[0008] In various embodiments of the invention, the disclosed methods are of use for a wide variety of applications for which nucleic acid sequencing is desired. Such applications include, but are not limited to, detecting the presence of substitutions, insertions and/or deletions in a nucleic acid template, such as chromosomal DNA or messenger RNA, detecting single nucleotide polymorphisms (SNPs), identifying unknown nucleic acids in a sample by sequence analysis, performing real-time PCR™ analysis and determining gene expression by sequencing. In particular embodiments, the BRC method may comprise a rolling circle method of nucleic acid replication.

[0009] In particular embodiments of the invention, pyrophosphate is generated by a reaction such as PCR, transcription and/or DNA replication. In preferred embodiments, sequence specific primers are used to limit replication to a particular template nucleic acid in the sample. The sequence specific primers are designed to not bind to other nucleic acids that may be present in the sample. The pyrophosphate may be reacted with APS in the presence of ATP sulfurylase to produce ATP and sulphate. The ATP may be reacted with oxygen and luciferin in the presence of luciferase to yield oxyluciferin, AMP and pyrophosphate. For each molecule of pyrophosphate that is cycled through BRC, a photon of light is emitted and one molecule of pyrophosphate is regenerated. Because of the relative kinetic rates of luciferase and ATP sulfurylase, a steady state is reached in which the concentrations of ATP and pyrophosphate and the level of photon output remain relatively constant over an extended period of time. The very high sensitivity of BRC is related in part to the integration of light output over time, in contrast to other methods that measure light output at a single time point or at a small number of fixed time points. The ability to vary the length of time over which photon integration occurs also contributes to the very high dynamic range for nucleic acid molecule quantification. The detection noise is also significantly reduced by increasing the length of integration.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0011] **FIG. 1** illustrates exemplary hypothetical waveforms for each of the bases adenine (A), guanine (G), cytosine (C) and thymine (T) that would be detected during DNA sequencing.

[0012] **FIG. 2** illustrates an exemplary hypothetical waveform generated for an exemplary DNA sequence TCTAGCTCAG (SEQ ID NO:1).

[0013] **FIG. 3** illustrates a noise-corrupted aggregate waveform obtained from a uniformly asynchronous reaction of 10^5 molecules of DNA with the exemplary sequence TCTAGCTCAG (SEQ ID NO:1).

[0014] **FIG. 4** illustrates a reconstructed waveform using the Wiener solution ($SNR_{perfect} = 40db$).

[0015] **FIG. 5** illustrates a reconstructed waveform using the Wiener solution ($SNR_{perfect} = 35db$).

[0016] **FIG. 6** illustrates a reconstructed waveform using the Wiener solution ($SNR_{perfect} = 30db$).

[0017] **FIG. 7** illustrates a reconstructed waveform using the Wiener solution ($SNR_{perfect} = 40db$ and $N = 10^6$).

[0018] **FIG. 8** shows an exemplary noise-corrupted aggregate waveform of 10^5 DNA molecules with Gaussian delay distribution.

[0019] **FIG. 9** illustrates an exemplary reconstructed waveform using the Wiener solution when the delay distribution is Gaussian ($SNR_{perfect} = 40db$).

[0020] **FIG. 10** shows an exemplary nanopore structure and current modulation due to blocking effect of a DNA strand.

[0021] **FIG. 11** shows an exemplary multi-aperture nanopore structure and current modulation due to the blocking effect of multiple identical DNA strands.

[0022] **FIG. 12** illustrates **(A)** background noise of the solution and **(B)** output of the sensor during nucleic acid polymerization.

[0023] **FIG. 13** illustrates an exemplary embodiment of the bioluminescence regenerative cycle (BRC) process.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Definitions

[0024] Terms that are not otherwise defined herein are used in accordance with their plain and ordinary meaning.

[0025] As used herein, "a" or "an" may mean one or more than one of an item.

[0026] As used herein, "luminescence" refers to the emission of light that does not derive energy from the temperature of the emitting body (i.e., emission of light other than incandescent light). "Luminescence" includes, but is not limited to, fluorescence, phosphorescence, thermoluminescence, chemiluminescence, electroluminescence and bioluminescence. "Luminescent" refers to an object that exhibits luminescence. In preferred embodiments, the light is in the visible spectrum. However, the present invention is not limited to visible light, but includes electromagnetic radiation of any frequency.

[0027] As used herein, the terms "analyte" and "template" mean any compound, molecule or aggregate of interest for sequence analysis. Non-limiting examples of templates include an oligonucleotide, polynucleotide, nucleic acid, peptide, polypeptide, protein, carbohydrate, polysaccharide, glycoprotein, lipid, hormone, growth factor, cytokine, receptor, antigen, allergen, antibody, toxin, poison, biowarfare agent, biohazardous agent, infectious agent, prion, or any other molecule without limitation, so long as it is capable of being sequenced. "Templates" are not limited to single molecules or atoms, but may also comprise complex aggregates, such as a virus, bacterium, *Salmonella*, *Streptococcus*, *Legionella*, *E. coli*, *Giardia*, *Cryptosporidium*, *Rickettsia*, spore, mold, yeast, algae, amoebae, dinoflagellate, unicellular organism, pathogen or cell. In certain embodiments, cells exhibiting a particular characteristic or disease state, such as a cancer cell, may be templates.

[0028] "Nucleic acid" means either DNA, RNA, single-stranded, double-stranded or triple stranded and any chemical modifications thereof. Virtually any modification of the nucleic acid is contemplated by this invention. "Nucleic acid" encompasses, but is not limited to, oligonucleotides and polynucleotides. Within the practice of the present invention, a "nucleic acid" may be of any length.

Model for Statistical Signal Processing

[0029] For a fixed signature signal $s(t)$ of duration T seconds, i.e,

$$s(t) \neq 0 \quad \text{for} \quad 0 \leq t \leq T$$

and

$$s(t) = 0 \quad \text{for} \quad t < 0, t > T$$

the random superposition of N such signatures immersed in noise may be observed. The observed signal may be described by

$$y(t) = \sum_{n=1}^{N} s(t - d_n) + v(t),$$

where d_n represents the random delay (time shift) for the n th ($n = 1,2,...,N$) signature sequence and where $v(t)$ represents the noise process. It is assumed that the observed signal starts at time t=0, so that all the delays are non-negative (i.e, $d_n \geq 0$).

[0030] In practice, continuous signals are rarely measured. Rather what are measured are the sampled values of the signal, obtained from sampling at a certain rate. With a sampling rate of R samples per second, the signature signal may be represented by the following sequence of length $L = RT + 1$

$$s_i = s(i/R), \quad i = 0,1,...,RT.$$

[0031] In this case, the sampled observation signal $y_i = y(i/R)$ is simply

$$y_i = \sum_{n=1}^{N} s_{i-k_n} + v_i, \tag{1}$$

where $v_i = v(i/R)$ represents the samples of the noise and where k_n represents the delay via the formula

$$k_n = \lfloor Rd_n \rfloor.$$

[0032] Equation 1 assumes that the sequence s_i is zero for $i \leq 0$. An important condition for the present analysis is that N is very large. In this case it is reasonable to consider a distribution for the delays d_n, or k_n. If N_j denotes the number of signature sequences that begin at time j, i.e., the number of n such that $k_n = j$, Equation 1 may be rewritten as

$$y_i = \sum_{j=0}^{D} N_j s_{i-j} + v_i, \tag{2}$$

where D represents the total duration of the delays. In other words, the delays j extend from $j = 0$ to $j = D$. Note, moreover, that

$$\sum_{j=0}^{D} N_j = N \tag{3}$$

and that the total duration of the observed signal is

$$D + RT + 1.$$

It is also possible to write the "convolution" in Equation 2 as

$$y_i = \sum_{j=0}^{RT} N_{i-j} s_j + v_j. \tag{4}$$

[0033] It is now possible to resolve the following problem. *Given the observations sequence y_i, satisfying Equation 2, or equivalently Equation 4, determine the unknown signature sequence s_i.* A standard assumption for the noise process v_i is that it is zero-mean, Gaussian and white, i.e., uncorrelated in time---although other types of noise models can also be dealt with, e.g., zero-mean Gaussian noise with a certain power spectral density function.

[0034] If the N_j in Equation 2 or Equation 4 are assumed to be known then we are simply confronted with an overdetermined system of linear equations in the unknowns s_i. To see this more explicitly, it is useful to rewrite Equation 2 in the following form

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{D+RT-1} \\ y_{D+RT} \end{bmatrix} = \begin{bmatrix} N_0 & & & \\ N_1 & N_0 & & \\ \vdots & N_1 & \ddots & \\ N_D & \vdots & \ddots & N_0 \\ & N_D & \vdots & N_1 \\ & & \ddots & \vdots \\ & & & N_D \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{RT-1} \\ s_{RT} \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{D+RT-1} \\ v_{D+RT} \end{bmatrix}. \quad (5)$$

[0035] With the N_j known, the coefficient matrix in Equation 5 is known. Therefore the unknown vector of s_i's can be readily computed via standard methods such as least-squares. The problem is that the N_j are not known. All that is observed is the sequence s_i. Therefore we are confronted with an equation where all the quantities on the right-hand-side (the N_j, s_i and v_i) are unknown. A natural question is whether in principle the desired s_i may be identified from Equation 5.

[0036] If it is assumed that the noise vector of v_i's is negligible, then Equation 5 is a system of $D+RT+1$ equations (the number of observations) in $D+RT+2$ unknowns ($D+1$ unknowns for the N_j and $RT+1$ unknowns for the s_i). Therefore, even in the noiseless case, it appears that there is an identifiability problem for there are more unknowns than equations. Of course, it is possible to use the equation $\sum_{j=0}^{D} N_j = N$ to get the number of equations and unknowns to match. However, with some very reasonable statistical assumptions it is possible to circumvent the idenitifiability problem altogether.

Statistical Assumptions: Exploiting Large N

[0037] A distinguishing feature of sequencing problems is that the number of DNA molecules, and hence signature sequences, N is extremely large. Therefore if something is known about the statistics of the delay distribution then it is possible to "estimate" the values of the N_j, and thereby the coefficient matrix in Equation 5. The statistics of the delay distribution is a *macroscopic* quantity, and so it is reasonable to assume that it is known. Moreover, being a

macroscopic quantity, it is also reasonable to assume that it may be controlled using an appropriate system design. This statistical knowledge can be used to estimate the N_j.

[0038] Uniform delay distribution: Assume that the delay distribution is uniform over D, the duration of the delays. In other words, the signature sequences are equally likely to begin anywhere in the interval $[0, D]$. This assumption is true in many applications and the sequencing system may be designed to exhibit a uniform delay distribution over D.

[0039] Using properties of the binomial distribution, each of the N_j will be random variables with mean and variance

$$\mu_N = EN_j = N/D \quad \text{and} \quad \sigma_N^2 = E(N_j - N/D)^2 = (1-1/D)N/D \tag{6}$$

where E denotes expectation. It can also be shown that the random variables N_j have cross-covariance:

$$C_{N_i N_j} = E(N_i - N/D)(N_j - N/D) = -N/D^2. \tag{7}$$

[0040] Equation 6 shows that as N grows larger the mean N/D becomes a better and better estimate of the actual value N_j. The ratio of the standard deviation of N_j to its mean is given by

$$\frac{\sigma_N}{\mu_N} = \sqrt{\frac{D-1}{N}},$$

which goes to zero as N goes to infinity, so that the estimate becomes more and more reliable with larger sample size. If we define the random variable $\tilde{N}_j = N_j - N/D$, Equation 5 may be rewritten as:

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{D+RT-1} \\ y_{D+RT} \end{bmatrix} = \frac{N}{D} \begin{bmatrix} 1 & & & & \\ 1 & 1 & & & \\ \vdots & 1 & \ddots & & \\ 1 & \vdots & \ddots & 1 & \\ & 1 & \vdots & 1 & \\ & & \ddots & \vdots & \\ & & & 1 & \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{RT-1} \\ s_{RT} \end{bmatrix} + \begin{bmatrix} \tilde{N}_0 & & & \\ \tilde{N}_1 & \tilde{N}_0 & & \\ \vdots & \tilde{N}_1 & \ddots & \\ \tilde{N}_D & \vdots & \ddots & \tilde{N}_0 \\ & \tilde{N}_D & \vdots & \tilde{N}_1 \\ & & \ddots & \vdots \\ & & & \tilde{N}_D \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{RT-1} \\ s_{RT} \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{D+RT-1} \\ v_{D+RT} \end{bmatrix}. \tag{8}$$

[0041] In Equation 8, the matrix coefficient in the first term is known. Although the second matrix coefficient is unknown its "energy" is less by a factor of N . To make this more precise, defining the last two terms in Equation 8 as an "equivalent" noise

$$\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{D+RT-1} \\ w_{D+RT} \end{bmatrix} = \begin{bmatrix} \tilde{N}_0 & & & & \\ \tilde{N}_1 & \tilde{N}_0 & & & \\ \vdots & \tilde{N}_1 & \ddots & & \\ \tilde{N}_D & \vdots & \ddots & \tilde{N}_0 & \\ & \tilde{N}_D & \vdots & \tilde{N}_1 & \\ & & \ddots & \vdots & \\ & & & \tilde{N}_D & \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{RT-1} \\ s_{RT} \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{D+RT-1} \\ v_{D+RT} \end{bmatrix}, \tag{9}$$

[0042] Using Equation 6 and Equation 7 it is straightforward to compute the covariance matrix of the equivalent noise. If the off-diagonal terms are ignored compared to the diagonals ones (from Equations 6 and 7 σ_N^2 is larger than $C_{N_i N_j}$ by a factor of D), then the covariance matrix can be written as

$$R_w = \begin{bmatrix} \sigma_v^2 + \frac{NP_s}{DRT} & & & & & & \\ & \sigma_v^2 + \frac{2NP_s}{DRT} & & & & & \\ & & \ddots & & & & \\ & & & \sigma_v^2 + \frac{NP_s}{D} & & & \\ & & & & \ddots & & \\ & & & & & \sigma_v^2 + \frac{2NP_s}{DRT} & \\ & & & & & & \sigma_v^2 + \frac{NP_s}{DRT} \end{bmatrix}$$

for $D > RT$ and

$$R_w = \begin{bmatrix} \sigma_v^2 + \frac{NP_s}{DRT} & & & & & \\ & \sigma_v^2 + \frac{2NP_s}{NRT} & & & & \\ & & \ddots & & & \\ & & & \sigma_v^2 + \frac{N(D+RT)}{2DRT} & & \\ & & & & \ddots & \\ & & & & & \sigma_v^2 + \frac{2NP_s}{DRT} \\ & & & & & & \sigma_v^2 + \frac{NP_s}{DRT} \end{bmatrix}$$

for $D < RT$, where the noise variance is defined as $Ev_i v_j = \sigma_v^2 \delta_{ij}$ and the signature signal energy is defined as

$$P_s = \sum_{i=0}^{RT} s_i^2 \,. \tag{10}$$

[0043] An important quantity is the "equivalent" signal-to-noise-ratio (SNR), which can be computed to be

$$SNR = \frac{SNR_{perfect}}{1 + \frac{D}{N} SNR_{perfect}}, \tag{11}$$

where

$$SNR_{perfect} = \frac{N^2 P_s}{D(D+RT)\sigma_v^2}, \tag{12}$$

is the SNR when we have *exact* knowledge of the N_j. As N goes to infinity, SNR approaches $SNR_{perfect}$. In other words, in the limit of large N, the system behaves as if the values of the N_j are known. Thus, the macroscopic statistical knowledge allows circumvention of the identifiability problem.

[0044] The Wiener solution: Now that all the relevant covariance matrices have been computed, it is straightforward to find the least-mean-squares estimate of the signature sequence. The solution is referred to as the *Wiener solution* and is given by

$$\begin{bmatrix} \hat{s}_0 \\ \hat{s}_1 \\ \vdots \\ \hat{s}_{RT-1} \\ \hat{s}_{RT} \end{bmatrix} = \frac{NP_s}{DRT}\Theta^*\left(R_w + \frac{N^2P_s}{D^2RT}\Theta\Theta^*\right)^{-1}\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{D+RT-1} \\ y_{D+RT} \end{bmatrix}, \tag{13}$$

where the $(D+RT+1)\times(RT+1)$ Toeplitz matrix Θ from Equation 8 is defined as

$$\Theta = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & 1 & \ddots & \\ 1 & \vdots & \ddots & 1 \\ & 1 & \vdots & 1 \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}.$$

[0045] The Wiener solution shown in Equation 13 requires computing the inverse of a $(D+RT+1)\times(D+RT+1)$ matrix. Due to the Toeplitz structure this can be done efficiently and in a numerically stable way. Examplary resolutions of the inverse matrix computation using the Wiener solution are provided below in the Examples section.

BRC Detection

[0046] Various embodiments of the invention concern novel methods for sequencing nucleic acid molecules without labeling of any template, capture or probe molecules. Such label free methods are advantageous with respect to sensitivity, expense and ease of use. The BRC methods involve the luminescent detection of pyrophosphate (PPi) molecules released from an enzyme-catalyzed reaction, such as RNA and/or DNA polymerization. As part of the technique, a bioluminescence regenerative cycle (BRC) is triggered by the release of inorganic pyrophosphate (PPi) from polymerization of a nucleic acid.

[0047] The regenerative cycle is illustrated in FIG. 13. It involves a first reaction of PPi with APS, catalyzed by ATP-sulfurylase enzyme, which results in the production of ATP and

inorganic sulphate. In a second reaction, luciferin and luciferase consume ATP as an energy source to generate light, AMP and oxyluciferin and to regenerate PPi (FIG. 13). Thus, after each BRC cycle, a quantum of light is generated for each molecule of PPi in solution, while the net concentration of ATP in solution remains relatively stable and is proportional to the initial concentration of PPi. In the course of the reactions, APS and luciferin are consumed and AMP and oxyluciferin are generated, while ATP sulfurylase and luciferase remain constant. The invention is not limited as to the type of luciferase used. Although certain disclosed embodiments utilized firefly luciferase, any luciferase known in the art may be used in the disclosed methods.

[0048] As a result of the BRC process, the photon emission rate remains steady and is a monotonic function of the amount of PPi in the initial mixture. For very low concentrations of PPi (10^{-8} M or less), the total number of photons generated in a fixed time interval is proportional to the number of PPi molecules. Where PPi is generated by the polymerase catalyzed replication of a template nucleic acid, the number of photons generated in a fixed time interval is proportional to the quantity of the template nucleic acid present in the sample.

[0049] In polymerase-catalyzed reactions, PPi molecules are generated when nucleotides (dNTPs) are incorporated with different rates into the nucleic acid chain. For each addition of a nucleotide, one PPi molecule is cleaved from the dNTP by the polymerase enzyme (e.g. Klenow) and released into the reaction buffer. The time delay from the PPi generation to the photon release in the BRC assay is limited by the kinetics of luciferase, which is a relatively slow enzyme compared to very fast catalytic reactions of both polymerase and ATP-sulfurylase. With current commercially enzymatic assays any fast change in the concentration of PPi molecules is less detectable in the light generation process. By decreasing the rate of polymerase activity (which is sequence dependant), polymerase kinetics become detectable in the output light intensity. Not all of the target molecules get polymerized at the exact same time, so the photon intensity is a function of all the individual polymerization steps which contain the sequence information.

[0050] The basic concept of enzymatic light generation from PPi molecules was introduced almost two decades ago (Nyren and Lundin, 1985; Nyren, *Anal. Biochem.* 167:235-238, 1987). Pyrophosphate based luminescence has been used for DNA sequencing (Ronaghi *et al.*, *Anal.*

Biochem. 242:84-89, 1996) and SNP detection (Nyren *et al.*, *Anal. Biochem.* 244:367-373, 1997).

[0051] The present disclosure provides for nucleic acid sequencing by BRC when all four nucleotide substrates are added simultaneously to the polymerization mixture. Compared to the sequential addition of individual nucleotides, polymerization with all four nucleotides results in a very complex signal, due to the time lag between polymerization of different strands of replicated nucleic acids. Such a signal must be deconvoluted, using the statistical methods discussed herein, in order to provide real time nucleic acid sequencing.

Enzymatic Bioluminescence Cycle

[0052] To generate photons from pyrophosphate, ATP-sulfurylase (Ronesto *et al.*, *Arch. Biochem. Biophys.* 290:66-78, 1994; Beynon *et al. Biochemistry*, 40, 14509-14517, 2001) is used to catalyze the transfer of the adenylyl group from APS to PPi, producing ATP and inorganic sulfate.

$$PPi + APS \longleftrightarrow ATP + SO_4^{-2}$$

[0053] Next, luciferase catalyzes the slow consumption of ATP, luciferin and oxygen to generate a single photon (λ_{max}=562nm, Q.E. ≈ 0.88) per ATP molecule, regenerating a molecule of PPi and producing AMP, CO_2 and oxyluciferin. (Brovko *et al.*, *Biochem.* (Moscow) 59:195-201, 1994)

$$ATP + Luciferin + O_2 \longrightarrow AMP + oxyluciferin + CO_2 + h\upsilon + PPi$$

[0054] Because the steady-state photon emission is proportional to the initial concentration of PPi, the presence of minute amounts of PPi produced by a polymerase or other reaction should result in a detectable shift in baseline luminescence, even in the presence of considerable amounts of noise. The number of photons generated over time by the BRC cycle can potentially be orders of magnitude higher than the initial number of PPi molecules, which makes the system extremely sensitive compared to other methods of nucleic acid sequencing.

Charge Perturbation Signature (CPS)

[0055] In other embodiments of the invention, charge perturbation signature (CPS) methods may be used for nucleic acid sequencing. CPS is based on measuring the variation of the net

charge of nucleic acid molecules (e.g. DNA), or other nano-biological entities, when they are exposed to different nucleotides or agents. In this method, primed single strand DNA molecules are immobilized on an electrode and placed in a solution containing polymerase. When nucleotides are added and extension occurs, the electrostatic response of a group of identical DNA molecules creates a unique waveform which can be used to recognize the pattern (for SNP or DNA sequencing), evaluate the mass (for gene expression), or localize the charge (for protein/protein interaction).

[0056] The chemistry used in charge-sequencing is simple because, unlike many other sequencing techniques that require multi-enzyme systems, only one enzyme is present in the solution (e.g. Klenow). The design of the sensor that measures the net charge variation is also relatively straight forward. The method of detection is real time because the charge perturbation of the immobilized DNA ions, which is used for reading the genetic code, occurs almost instantaneously once the electrode is placed in the solution containing the polymerase and nucleotides (less than 1 second for a DNA fragment of 100 base pairs).

[0057] In the CPS method, the overall electrical pattern generated by the specific molecular binding is the sum of the individual patterns for all template molecules undergoing replication. As discussed above, the replication of different copies of the same template strand may not be synchronized. The rate of polymerization of nucleotides is intrinsically different for different nucleotides. Thus the charge accumulation detected by CPS will also be different, depending on which nucleotides are undergoing incorporation. Given that, the individual electrical signature of a "single" molecule may be deconvoluted by the statistical analysis methods disclosed herein.

[0058] A prototype nucleic acid sequencing apparatus utilizing CPS techniques has been built and tested. The results (not shown) demonstrated that the signal created by polymerization is detectable and can be easily acquired using commercial electronic components. Without extensive optimization, the sensitivity of the prototype was comparable to that of an optical detection system. Thus, charge-sequencing platforms should provide enhanced sensitivity compared to presently available sequencing apparatus. The advantages of charge-sequencing include real time detection, simple chemistry, suitability for integration on chip based electrical detection systems (e.g. CMOS) and suitability for different applications in genomics and proteomics, including but not limited to nucleic acid sequencing.

Nucleic Acids

[0059] Samples comprising nucleic acids may be prepared by any technique known in the art. In certain embodiments, analysis may be performed on crude sample extracts, containing complex mixtures of nucleic acids, proteins, lipids, polysaccharides and other compounds. Such samples are likely to contain contaminants that could potentially interfere with the sequencing process, although the use of primers specific for one or more target nucleic acid sequences may decrease or even eliminate the problem of trying to sequence multiple templates of different sequence simultaneously. In preferred embodiments, specific target nucleic acids may be partially or fully separated from other sample constituents before initiating the BRC analysis.

[0060] Methods for partially or fully purifying DNA and/or RNA from complex mixtures, such as cell homogenates or extracts, are well known in the art. (See, *e.g.*, Guide to Molecular Cloning Techniques, eds. Berger and Kimmel, Academic Press, New York, NY, 1987; Molecular Cloning: A Laboratory Manual, 2nd Ed., eds. Sambrook, Fritsch and Maniatis, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989). Generally, cells, tissues or other source material containing nucleic acids are first homogenized, for example by freezing in liquid nitrogen followed by grinding in a mortar and pestle. Certain tissues may be homogenized using a Waring blender, Virtis homogenizer, Dounce homogenizer or other homogenizer. Crude homogenates may be extracted with detergents, such as sodium dodecyl sulphate (SDS), Triton X-100, CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate), octylglucoside or other detergents known in the art. As is well known, nuclease inhibitors such as RNase or DNase inhibitors may be added to prevent degradation of template nucleic acids.

[0061] Extraction may also be performed with chaotrophic agents such as guanidinium isothiocyanate, or organic solvents such as phenol. In some embodiments, protease treatment, for example with proteinase K, may be used to degrade cell proteins. Particulate contaminants may be removed by centrifugation or ultracentrifugation. Dialysis against aqueous buffer of low ionic strength may be of use to remove salts or other soluble contaminants. Nucleic acids may be precipitated by addition of ethanol at -20°C, or by addition of sodium acetate (pH 6.5, about 0.3 M) and 0.8 volumes of 2-propanol. Precipitated nucleic acids may be collected by centrifugation or, for chromosomal DNA, by spooling the precipitated DNA on a glass pipet or other probe. The skilled artisan will realize that the procedures listed above are exemplary only

and that many variations may be used, depending on the particular type of nucleic acid to be analyzed.

[0062] In certain embodiments, the nucleic acids to be analyzed may be naturally occurring DNA or RNA molecules. Virtually any naturally occurring nucleic acid may be analyzed by the disclosed methods including, without limit, chromosomal, mitochondrial or chloroplast DNA or ribosomal, transfer, heterogeneous nuclear or messenger RNA. Nucleic acids may be obtained from either prokaryotic or eukaryotic sources by standard methods known in the art. Alternatively, nucleic acids of interest may be prepared artificially, for example by PCR™ or other known amplification processes or by preparation of libraries such as BAC, YAC, cosmid, plasmid or phage libraries containing nucleic acid inserts. (See, *e.g.*, Berger and Kimmel, 1987; Sambrook *et al.*, 1989.) The source of the nucleic acid is unimportant for purposes of performing sequence analysis and it is contemplated within the scope of the invention that nucleic acids from virtually any source may be analyzed using the methods disclosed herein.

Methods of Immobilization

[0063] In various embodiments, the nucleic acids to be analyzed may be attached to a solid surface (or immobilized). Immobilization of nucleic acids may be achieved by a variety of methods involving either non-covalent or covalent attachment between the nucleic acid and the surface. In an exemplary embodiment, immobilization may be achieved by coating a surface with streptavidin or avidin and the subsequent attachment of a biotinylated polynucleotide (Holmstrom *et al.*, *Anal. Biochem.* 209:278-283, 1993). Immobilization may also occur by coating a silicon, glass or other surface with poly-L-Lys (lysine), followed by covalent attachment of either amino- or sulfhydryl-modified nucleic acids using bifunctional crosslinking reagents (Running *et al.*, *BioTechniques* 8:276-277, 1990; Newton *et al.*, *Nucleic Acids Res.* 21:1155-62, 1993). Amine residues may be introduced onto a surface through the use of aminosilane for cross-linking.

[0064] Immobilization may take place by direct covalent attachment of 5'-phosphorylated nucleic acids to chemically modified surfaces (Rasmussen *et al.*, *Anal. Biochem.* 198:138-142, 1991). The covalent bond between the nucleic acid and the surface is formed by condensation with a water-soluble carbodiimide. This method facilitates a predominantly 5'-attachment of the nucleic acids via their 5'-phosphates.

[0065] DNA is commonly bound to glass by first silanizing the glass surface, then activating with carbodiimide or glutaraldehyde. Alternative procedures may use reagents such as 3-glycidoxypropyltrimethoxysilane (GOP) or aminopropyltrimethoxysilane (APTS) with DNA linked *via* amino linkers incorporated either at the 3′ or 5′ end of the molecule. DNA may be bound directly to membrane surfaces using ultraviolet radiation. Other non-limiting examples of immobilization techniques for nucleic acids are disclosed in U.S. Patent Nos. 5,610,287, 5,776,674 and 6,225,068.

[0066] The type of surface to be used for immobilization of the nucleic acid is not limiting. In various embodiments, the immobilization surface may be magnetic beads, non-magnetic beads, a planar surface, or any other conformation of solid surface comprising almost any material, so long as the material is sufficiently durable and inert to allow the BRC process to occur. Non-limiting examples of surfaces that may be used include glass, silica, silicate, PDMS, silver or other metal coated surfaces, nitrocellulose, nylon, activated quartz, activated glass, polyvinylidene difluoride (PVDF), polystyrene, polyacrylamide, other polymers such as poly(vinyl chloride), poly(methyl methacrylate) or poly(dimethyl siloxane), and photopolymers which contain photoreactive species such as nitrenes, carbenes and ketyl radicals capable of forming covalent links with nucleic acids (See U.S. Pat. Nos. 5,405,766 and 5,986,076).

[0067] Bifunctional cross-linking reagents may be of use in various embodiments, such as attaching a nucleic acid to a surface. The bifunctional cross-linking reagents can be divided according to the specificity of their functional groups, *e.g.*, amino, guanidino, indole, or carboxyl specific groups. Of these, reagents directed to free amino groups are popular because of their commercial availability, ease of synthesis and the mild reaction conditions under which they can be applied. Exemplary methods for cross-linking molecules are disclosed in U.S. Patent Nos. 5,603,872 and 5,401,511. Cross-linking reagents include glutaraldehyde (GAD), bifunctional oxirane (OXR), ethylene glycol diglycidyl ether (EGDE), and carbodiimides, such as 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC).

Polymerases

[0068] In certain embodiments, the disclosed methods may involve binding of a DNA polymerase to a primer molecule and the catalyzed addition of nucleotide precursors to the 3' end of a primer. Non-limiting examples of polymerases of potential use include DNA

polymerases, RNA polymerases, reverse transcriptases, and RNA-dependent RNA polymerases. The differences between these polymerases in terms of their requirement or lack of requirement for primers and promoter sequences are known in the art.

[0069] Non-limiting examples of polymerases that may be of use include *Thermatoga maritima* DNA polymerase, AmplitaqFS™ DNA polymerase, Taquenase™ DNA polymerase, ThermoSequenase™, Taq DNA polymerase, Qbeta™ replicase, T4 DNA polymerase, *Thermus thermophilus* DNA polymerase, RNA-dependent RNA polymerase and SP6 RNA polymerase. Commercially available polymerases including Pwo DNA Polymerase from Boehringer Mannheim Biochemicals (Indianapolis, IN); Bst Polymerase from Bio-Rad Laboratories (Hercules, CA); IsoTherm™ DNA Polymerase from Epicentre Technologies (Madison, WI); Moloney Murine Leukemia Virus Reverse Transcriptase, *Pfu* DNA Polymerase, Avian Myeloblastosis Virus Reverse Transcriptase, *Thermus flavus (Tfl)* DNA Polymerase and *Thermococcus litoralis (Tli)* DNA Polymerase from Promega (Madison, WI); RAV2 Reverse Transcriptase, HIV-1 Reverse Transcriptase, T7 RNA Polymerase, T3 RNA Polymerase, SP6 RNA Polymerase, RNA Polymerase E. coli, *Thermus aquaticus* DNA Polymerase, T7 DNA Polymerase +/- 3'→5' exonuclease, Klenow Fragment of DNA Polymerase I, Thermus 'ubiquitous' DNA Polymerase, and DNA polymerase I from Amersham Pharmacia Biotech (Piscataway, NJ).

[0070] As is known in the art, various polymerases have an endogenous 3'-5' exonuclease activity that may be used for proof-reading newly incorporated nucleotides. Because a molecule of pyrophosphate is generated for each nucleotide incorporated into a growing chain, regardless of whether or not it is subsequently removed, in certain embodiments of the invention it may be preferred to use polymerases that lack exonuclease or proof-reading activity. Methods of using polymerases and compositions suitable for use in such methods are well known in the art (*e.g.*, Berger and Kimmel, 1987; Sambrook *et al.*, 1989).

Primers

[0071] Where primers are required to initiate polymerase activity, they may be obtained by any method known in the art. Generally, primers are between ten and twenty bases in length, although longer primers may be employed. In certain embodiments, primers are designed to be exactly complementary in sequence to a known portion of a template nucleic acid, preferably at

or close to the 3' end of the template nucleic acid. Methods for synthesis of primers of any sequence are known, for example using an automated nucleic acid synthesizer employing phosphoramidite chemistry. Such instruments may be obtained from commercial sources, such as Applied Biosystems (Foster City, CA) or Millipore Corp. (Bedford ,MA).

Template Dependent Amplification Methods

[0072] A number of template dependent processes are available to amplify the sequences present in a given nucleic acid template. One of the best known amplification methods is polymerase chain reaction (referred to as PCR) which is disclosed in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, and in Innis *et al.*, *PCR Protocols*, Academic Press, Inc., San Diego CA, 1990.

[0073] A reverse transcriptase PCR amplification procedure may be performed in order to amplify mRNA. Methods of reverse transcribing RNA into cDNA are well known (Sambrook *et al.*, 1989). Alternative methods for reverse transcription utilize thermostable DNA polymerases are well known in the art.

[0074] Another method for amplification is ligase chain reaction ("LCR"). In LCR, two complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe pairs. U.S. Patent 4,883,750 discloses a method similar to LCR for binding probe pairs to a target sequence.

[0075] Qbeta Replicase may also be used as another amplification method. In this method, a replicative sequence of RNA which has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence which may then be detected.

[0076] An isothermal amplification method, in which restriction endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[alpha-thio]-triphosphates in one strand of a restriction site may also be useful in the amplification of nucleic acids (*e.g.*, Walker *et al.*, *Proc. Nat'l Acad. Sci. USA* **89**:392-396, 1992).

[0077] Strand Displacement Amplification (SDA) is another method of carrying out isothermal amplification of nucleic acids, involving multiple rounds of strand displacement and synthesis. A similar method, called Repair Chain Reaction (RCR), involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction in which only two of the four bases are present. The other two bases may be added as biotinylated derivatives for easy detection. A similar approach is used in SDA. Target specific sequences may also be detected using a cyclic probe reaction (CPR). In CPR, a probe having 3' and 5' sequences of non-specific DNA and a middle sequence of specific RNA is hybridized to DNA which is present in a sample. Upon hybridization, the reaction is treated with RNase H, and the products of the probe identified as distinctive products which are released after digestion. The original template is annealed to another cycling probe and the reaction is repeated.

[0078] Other known nucleic acid amplification procedures include transcription-based amplification systems (TAS), including nucleic acid sequence based amplification (NASBA) and 3SR (*e.g.*, Kwoh *et al.*, *Proc. Nat'l Acad. Sci. USA* **86**:1173, 1989). In NASBA, the nucleic acids may be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. The amplification techniques involve annealing a primer that has target specific sequences. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target specific primer, followed by polymerization. The double-stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNAs are reverse transcribed into double stranded DNA, and transcribed once against with a polymerase such as T7 or SP6.

[0079] Another nucleic acid amplification process involves cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). The ssRNA is a first template for a first primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from the resulting DNA:RNA duplex by the action RNase H. The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to the template. This primer is then extended by DNA

polymerase (e.g., Klenow fragment), resulting in a double-stranded DNA ("dsDNA") molecule, having a sequence identical to the original RNA between the primers and having additionally, at one end, a promoter sequence. This promoter sequence may be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies may then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification may be done isothermally without addition of enzymes at each cycle.

[0080] Other exemplary amplification methods include "race" and "one-sided PCR." (Frohman, M.A., In: *PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS*, Academic Press, NY, 1990; Ohara *et al.*, *Proc. Nat'l Acad. Sci. USA*, **86**:5673-5677, 1989). Methods based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, may also be used in amplification (Wu *et al.*, *Genomics* 4:560, 1989). The disclosed methods are not limiting and any amplification method known in the art may be used in the practice of the claimed subject matter.

Detectors

[0081] In various embodiments of the invention, photons generated by BRC may be quantified using a detector, such as a charge coupled device (CCD). Other exemplary detectors include photodiodes, avalanche photodiodes, photomultiplier tubes, multianode photomultiplier tubes, phototransistors, vacuum photodiodes, silicon photodiodes, and CCD cameras.

[0082] In certain embodiments of the invention, a highly sensitive cooled CCD detector may be used. The cooled CCD detector has a probability of single-photon detection of up to 80%, a high spatial resolution pixel size (5 microns), and sensitivity in the visible through near infrared spectra. (Sheppard, Confocal Microscopy: Basic Principles and System Performance in: Multidimensional Microscopy, P.C. Cheng *et al.* eds., Springer-Verlag, New York, NY pp. 1-51, 1994.) In another embodiment of the invention, a coiled image-intensified coupling device (ICCD) may be used as a photodetector that approaches single-photon counting levels (U.S. Pat. No. 6,147,198). A small number of photons triggers an avalanche of electrons that impinge on a phosphor screen, producing an illuminated image. This phosphor image is sensed by a CCD chip region attached to an amplifier through a fiber optic coupler.

[0083] In some embodiments of the invention, an avalanche photodiode (APD) may be made to detect low light levels. The APD process uses photodiode arrays for electron multiplication effects (U.S. Pat. No. 6,197,503). The invention is not limited to the disclosed embodiments and it is contemplated that any light detector known in the art that is capable of accumulating photons over a time interval may be used in the disclosed methods and apparatus.

[0084] In all of the above embodiments the generated photons from the sample can either reach the detector directly or be guided and/or focused onto the detector by a secondary system such as a number of lenses, reflecting mirror systems, optical waveguides and optical fibers or a combination of those.

Nanopores

[0085] In certain embodiments of the invention, nanopores may be used to characterize a group of identical nucleic acids or even individual DNA or RNA molecules, using channel recording techniques to detect electrical signals generated by the movement of nucleic acids through the nanopores (FIG. 10). A nanopore may be a protein channel in a lipid bilayer or an extremely small isolated *'hole'* in a thin, solid-state membrane. For a nanopore to be useful as a single molecule detector, its diameter must not be much larger than the size of the molecule to be detected. When a single molecule enters a nanopore in an insulating membrane, it causes changes in the nanopore's electrical properties that are readily detected with known electronic devices and circuits.

[0086] Nanopores formed from alpha-hemolysin in lipid bilayers have been used to demonstrate the feasibility of this approach (not shown). The principles underlying this technology are exemplified in a device with a membrane separating two buffer filled chambers (FIG. 10). A channel in a membrane connects two ionic solution-filled compartments, designated N (negative) and P (positive) (FIG. 10). Ions flow through the open channel in response to a voltage applied across the membrane. When DNA, which is negatively charged, is added to the N compartment, the negatively charge DNA molecules are driven, one at a time, into and through the channel. Ionic current is reduced during the time the channel is occupied or "blocked" by the DNA.

[0087] The voltage applied across the bilayer (e.g. hundreds of mV) produces a current of ions that flows through the channel. After adding single stranded DNA molecules, transient

reductions of the current through the channel are evident as each of several DNA molecules are drawn through the channel. Because the pore in the channel is large enough to admit only a single strand of DNA, the duration of the ionic current block effectively measures the length of the DNA molecule. The extent to which ionic flow is reduced reflects the physical and chemical properties of the nucleotides in the polymer, and current modulation represents the structure and quantity of the molecule (e.g. DNA sequencing).

[0088] The amplitude of current modulation for a single nucleic acid strand moving through a single nanopore is relatively low. Measurable current modulation may be provided by using multiple nucleic acids moving through multiple nanopores. As discussed above, the complex signal generated by such a system would represent a cumulative signal produced by multiple copies of a template nucleic acid, which may move through the nanopores in a temporally offset pattern. The signal produced by this kind of system (FIG. 11) represents an electrical *signature*, which is the sum of all single molecules passing through different pores. Such complex signals may be deconvoluted using the disclosed statistical analysis methods to produce nucleic acid sequence data.

Fabrication of Nanopore

[0089] Fabrication of nanopores, individually or in arrays, may utilize any technique known in the art for nanoscale manufacturing. In certain embodiments of the invention, nanopores may be constructed on a solid-state matrix using known nanolithography methods, including but not limited to chemical vapor deposition, electrochemical deposition, chemical deposition, electroplating, thermal diffusion and evaporation, physical vapor deposition, sol-gel deposition, focused electron beam, focused ion beam, molecular beam epitaxy, dip-pen nanolithography, reactive-ion beam etching, chemically assisted ion beam etching, microwave assisted plasma etching, electro-oxidation, scanning probe methods, chemical etching, laser ablation, or any other method known in the art (*E.g.*, U.S. Patent No. 6,146,227).

[0090] In certain embodiments of the invention, channels or grooves may be etched into a semiconductor surface by various techniques known in the art including, but not limited to, methodologies using an STM/AFM tip in an oxide etching solution. After channels are formed, two semiconductor surfaces may be opposed to create one or more nanopores that penetrate the semiconductor. In other embodiments of the invention, STM tip methodologies may be used to create nanopores and other nanostructures using techniques known in the art. In alternative

embodiments of the invention, scanning probes, chemical etching techniques, and/or micromachining may be used to cut micrometer-dimensioned or nanometer-dimensioned channels, grooves or holes in a semiconductor substrate.

[0091] Alternatively, nanopores may be made using a high-throughput electron-beam lithography system (*e.g.*, http://www.mdatechnology.net/ techsearch.asp?articleid=510). Electron-beam lithography may be used to write features as small as 5 nm on silicon chips. Sensitive resists, such as polymethyl-methacrylate, coated on silicon surfaces may be patterned without use of a mask. The electron-beam array may combine a field emitter cluster with a microchannel amplifier to increase the stability of the electron beam, allowing operation at low currents. In some embodiments of the invention, the SoftMask™ control system may be used to control electron-beam lithography of nanoscale features on a semiconductor chip substrate.

[0092] In alternative embodiments of the invention, nanopores may be produced using focused atom lasers (*e.g.*, Bloch *et al.*, "Optics with an atom laser 210 beam," *Phys. Rev. Lett.* 87:123-321,1). Focused atom lasers may be used for lithography, much like standard lasers or focused electron beams. Such techniques are capable of producing micron scale or even nanoscale structures on a chip. In other alternative embodiments of the invention, dip-pen nanolithography may be used to form (*e.g.*, Ivanisevic *et al.*, "Dip-Pen Nanolithography on Semiconductor Surfaces," *J. Am. Chem. Soc.*, 123: 7887-7889,1). Dip-pen nanolithograpy uses AFM techniques to deposit molecules on surfaces, such as silicon chips. Features as small as 15 nm in size may be formed, with spatial resolution of 10 nm. Nanoscale pores may be formed by using dip-pen nanolithography in combination with regular photolithography techniques.

[0093] In other embodiments of the invention, ion-beam lithography may be used to create on a chip (*e.g.*, Siegel, "Ion Beam Lithography," VLSI Electronics, Microstructure Science, Vol. 16, Einspruch and Watts Eds., Academic Press, New York, 1987). A finely focused ion beam may be used to write nanoscale features directly on a layer of resist without use of a mask. Alternatively, broad ion beams may be used in combination with masks to form features as small as 100 nm in scale. Chemical etching, for example, with hydrofluoric acid, is used to remove exposed silicon or other chip material that is not protected by resist. The disclosed methods are not limiting, and nanopores may be formed by any method known in the art.

EXAMPLES

Example 1: Statistical Data Analysis for Nucleic Acid Sequencing

[0094] The signals considered in the following Example are in no way meant to model the actual signals obtained from sequencing DNA molecules. Instead, they provide an exemplary construct for demonstrating the use of the subject statistical analysis techniques to deconvolute complex signals. It is expected that the signals detected during sequencing will differ, depending on the measurement technique (nanopore, charge perturbation, bioluminescence regeneration, etc.) being employed. However, it is also expected that whatever detection system is used, the signals generated by incorporation of different types of nucleotide (A, T, G, C) will be distinguishable from each other.

[0095] An exemplary waveform generated by the reaction of each of the bases A, C, G and T is given in FIG. 1. For purposes of illustration, the sequence of the subject nucleic acid is given as TCTAGCTCAG (SEQ ID NO:1). The resulting waveform is shown in FIG. 2. The hypothetical sampling yields the duration of the signal as $RT = 160$.

[0096] If it is assumed that there are 10^5 such molecules in a sample and that they all react in a totally asynchronous fashion, with a uniformly distributed delay over the interval [0,400], then the noise-corrupted aggregate signal for a random run (i.e., what would be measured) is shown in FIG. 3, where $D = 400$ and the total duration of the observed signal is $D + RT = 560$.

[0097] Next assume that the signal-to-noise-ratio is given by $SNR_{perfect} = 40db$. Due to the lack of knowledge of the N_j, the effective SNR computed from Equation 11 is $SNR = 23.8db$. Using the Wiener solution shown in Equation 13, the recovered signature sequence is shown in FIG. 4. As can be seen, the signature signal has been successfully recovered and simple techniques, such as matched filtering, can be employed to reconstruct the sequence TCTAGCTCAG (SEQ ID NO:1).

[0098] The performance of the system depends on the SNR. If the SNR is decreased to $SNR_{perfect} = 35db$ and $SNR = 23.6db$ then, as shown in FIG. 5, the performance degrades, although the sequence is still recoverable. However, if the SNR is further decreased to $SNR_{perfect} = 30db$ and $SNR = 23db$ then, as shown in FIG. 6, the sequence is barely recoverable.

[0099] As discussed above, increasing the number of template molecules in the sample improves the performance of the system. Increasing the number of molecules to 10^6

automatically makes the signal strength 100 times stronger. To have a fair comparison the SNR is set to $SNR_{perfect} = 40db$ increasing the noise by 100-fold. With the new value of N we obtain $SNR = 28.5db$. The resulting recovered signature signal is shown in FIG. 7, which shows a clear improvement over FIG. 4.

Example 2: Non-Uniform Delay Distribution

[0100] Example 1 shows that using large template numbers and the Wiener solution, it is possible to recover the signature sequence from the noise-corrupted aggregate signal of its random shifts. However, the values of SNR to make this possible appear to be rather high. In the following discussion it is assumed that the duration of the delays is longer than the duration of each signature sequence, i.e., $D > RT$. This appears to be more appropriate for actual sequencing applications than $D \leq RT$.

[0101] Close inspection of Equation 8 shows that, of the $D + RT + 1$ equations, only $2RT + 1$ are distinct. These correspond to the first RT equations (coming from the "lower triangular" portion of the "diamond-shaped" matrix Θ) and the last RT equations (coming from the "upper triangular" portion of Θ). The $D - RT + 1$ center equations are all identical since the corresponding rows of Θ are all-one row-vectors. Thus, these $D - RT + 1$ equations all amount to a single equation giving information about

$$s_0 + s_1 + \ldots + s_{RT-1} + s_{RT},$$

i.e., the mean of the signature sequence. Physically, this means that information on the signature sequence can only be found in the *rising transient,* consisting of the first RT samples, of the observations signal and in its *falling transient,* consisting of the last RT samples. The middle $D - RT + 1$ samples, corresponding to the *steady-state* of the observed signal, contain no information on the signature sequence other than its mean. (These three phases can be clearly seen in FIG. 3, for example.) Of the N DNA molecules only those that begin (end) their reaction during the first (last) RT samples provide information on the signature sequence. The information carried by the remaining DNA molecules gets "averaged-out" during the steady-state of the observations signal.

[0102] It may be possible to improve performance by obtaining information from *all* the DNA molecules and ensuring that the observations signal never goes to a steady-state. The most

straightforward way to guarantee this is to have a *non-uniform* delay distribution. Thus, assume that the delay distribution is given by the probability sequence p_i, where

p_i = probability that the n th delay for an arbitrary n is $k_n = i$, for $i = 0,1,...,D$.

[0103] Being a probability sequence,

$$\sum_{i=0}^{D} p_i = 1.$$

[0104] Under this assumption the N_j are random variables with

$$\mu_{N_j} = EN_j = Np_j \quad \text{and} \quad \sigma_{N_j}^2 = E(N_j - Np_j)^2 = Np_j(1-p_j) \tag{14}$$

and

$$C_{N_i N_j} = E(N_i - Np_i)(N_j - Np_j) = -Np_i p_j. \tag{15}$$

[0105] Equations 14 and 15 reduce to Equations 6 and 7 when the distribution is uniform and $p_i = 1/D$ for all i. It follows that

$$\frac{\sigma_{N_j}}{\mu_{N_j}} = \sqrt{\frac{1-p_j}{Np_j}},$$

so that as N goes to infinity, we may replace N_j by its mean. Moreover, Equation 8 is replaced by

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{D+RT-1} \\ y_{D+RT} \end{bmatrix} = N \begin{bmatrix} p_0 & & & & \\ p_1 & p_0 & & & \\ \vdots & p_1 & \ddots & & \\ p_D & \vdots & \ddots & p_0 & \\ & p_D & \vdots & p_1 & \\ & & \ddots & \vdots & \\ & & & p_D & \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{RT-1} \\ s_{RT} \end{bmatrix} + \begin{bmatrix} \tilde{N}_0 & & & & \\ \tilde{N}_1 & \tilde{N}_0 & & & \\ \vdots & \tilde{N}_1 & \ddots & & \\ \tilde{N}_D & \vdots & \ddots & \tilde{N}_0 & \\ & \tilde{N}_D & \vdots & \tilde{N}_1 & \\ & & \ddots & \vdots & \\ & & & \tilde{N}_D & \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_{RT-1} \\ s_{RT} \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{D+RT-1} \\ v_{D+RT} \end{bmatrix} \tag{16}$$

[0106] As before R_w, the covariance matrix of the equivalent noise, may be computed. The "equivalent" SNR can be computed to be

$$SNR = \frac{SNR_{perfect}}{1 + \frac{\sum_{i=0}^{D} p_i^2}{N \sum_{i=0}^{D} p_i (1 - p_i)} SNR_{perfect}}, \tag{17}$$

where

$$SNR_{perfect} = \frac{N^2 P_s \sum_{i=0}^{D} p_i^2}{(D + RT)\sigma_v^2}. \tag{18}$$

[0107] The Wiener solution is still given by Equation 13. The only difference is that

$$\Theta = D \begin{bmatrix} p_0 & & & \\ p_1 & p_0 & & \\ \vdots & p_1 & \ddots & \\ p_D & \vdots & \ddots & p_0 \\ & p_D & \vdots & p_1 \\ & & \ddots & \vdots \\ & & & p_D \end{bmatrix}.$$

[0108] As an example of such a system, assume that the delay distribution has a Gaussian profile. Then, assuming that $N = 10^5$, FIG. 3 for the aggregate signal is replaced by FIG. 8, which demonstrates the absence of a steady-state. The recovered signature sequence using the Wiener solution is given in FIG. 9. The performance is comparable to that of the uniform delay case of FIG. 4.

Example 3: Engineering the delay distribution

[0109] The delay distribution has an effect on the number of DNA molecules that contribute to the recovery of the signature sequence. Although the particular Gaussian delay distribution considered in Example 2 did not appear to offer an improvement over a uniform delay distribution, there are likely to be delay distributions that provide improved results compared to uniform delay distribution.

Iteratively estimating the N_j

[0110] In the algorithm described above, the N_j were estimated using statistical assumptions on the delay distribution. Once the signature sequence has been estimated based on these statistical assumptions, it is possible to rewrite Equation 4 as Equation 19 below, where the roles

of the N_j and s_i have been exchanged. With the s_i estimated, the coefficient matrix in Equation 19 is known and the N_j may be estimated using the Wiener solution. Returning to Equation 5, the s_i, may be re-estimated and so on. Several iterations of this process should significantly improve the performance of the system.

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{D+RT-1} \\ y_{D+RT} \end{bmatrix} = \begin{bmatrix} s_0 & & & & \\ s_1 & s_0 & & & \\ \vdots & s_1 & \ddots & & \\ s_{RT} & \vdots & \ddots & s_0 & \\ & s_{RT} & \vdots & s_1 & \\ & & \ddots & \vdots & \\ & & & s_{RT} & \end{bmatrix} \begin{bmatrix} N_0 \\ N_1 \\ \vdots \\ N_{D-1} \\ N_D \end{bmatrix} + \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{D+RT-1} \\ v_{D+RT} \end{bmatrix}, \tag{19}$$

Nonlinear techniques:

[0111] The Wiener solution is essentially a linear technique, in the sense that the solution is linear in the observed signal y_i. The solution also requires no *a priori* knowledge of the signature sequence, i.e., that it is composed of components generated by the different base pairs A, C, G and T. This requires no assumptions on the signature sequence and so may be used to discover what the different signals generated by the A, G, C and T bases are. However, if the nature of the signature sequence is known, then it should be possible to improve the estimation. For example, if it is known that the signature sequence is composed of A, C, G and T components, each with a known response, then *nonlinear* techniques exist that can exploit this and improve performance. Among these are decision-feedback and maximum-likelihood techniques, the latter of which can be efficiently implemented using the Viterbi algorithm or sphere decoding.

Exploiting the spectral properties of the signature sequence:

[0112] In the Wiener solution it is not assumed that any information on the spectrum of the signature sequence is available. For example, it is reasonable to assume that the signature sequence is relatively smooth (and so devoid of high frequency components). However, the reconstructed sequences in the above Examples exhibit rapid time variations, indicating the presence of high frequency components. Therefore appropriate filtering of the recovered sequence should also improve the performance of the system.

More Complex Models

[0113] The Examples above assume that there is a *unique* signature sequence obtained from the DNA molecules that is repeatedly generated, with different delays. In other words, it is assumed that all molecules react in an identical fashion. However, it is possible that the actual signals may vary, depending on their context. For example, it is possible that base A may react differently if it is preceded by an A, C, G or T. It is known that poly CG sequences may polymerize differently than more random sequences. It is plausible that the time between the signal obtained from *each* of the bases may be random. In this case, rather than having a unique signature sequence, there will be a *distribution* of signature sequences. In certain alternative embodiments may utilize a stochastic model for the generation of the signature sequences, such as a hidden Markov model. Many algorithms for the estimation of the state of hidden Markov models are known in the art and may be used in the practice of the claimed methods.

Example 4: BRC Assay

Sample Preparation

[0114] Total RNA extracts may be obtained from blood, tissues or cell lines using commercially available kits (*e.g.*, Ambion, Austin, TX; Qiagen, Valencia, CA; Promega, Madison, WI). cDNA may be synthesized using a SuperScript™ or other commercial kit (Invitrogen Life Technologies, Austin, TX). Where preferred, polyadenylated mRNA may be purified by oligo(dT) column chromatography or other known methods.

[0115] In an exemplary embodiment, first strand cDNA synthesis employed an RNA/primer mixture containing 5 μl total RNA and 1 μl of 0.5 μg/μl oligo(dT) random primer or gene specific primer, incubated at 70°C for 10 min and then placed on ice for at least 1 min. A reaction mixture containing 2 μl 10X buffer (0.1 M Tris-Acetate pH 7.75, 5 mM EDTA, 50 mM Mg-acetate, 2 mM kinase free dNTP and 0.1 M dithiothreitol) in which dATP was replaced with α-thio dATP was added to the RNA/primer mixture, mixed gently, collected by brief centrifugation and then incubated at 42°C for 5 min. After addition of 200 U of SuperScript II reverse transcriptase, the tube was incubated at 40°C for 15 min. The reaction was terminated by heating at 70°C for 15 min and then chilling on ice. The dNTP used in cDNA synthesis should be kinase free. In preferred embodiments dATP is replaced with alpha-thio dATP or analogs that are not good substrates for luciferase.

[0116] An aliquot of synthesized cDNA was added to 50 μl of reaction mixture (see Ronaghi *et al., Anal. Biochem.* 242:84-89, 1996 with modifications) containing 250 ng luciferase (Promega, Madison, WI), 50 mU ATP sulfurylase (Sigma Chemical Co., St. Louis, MO), 2 mM dithiothreitol, 100 mM Tris-Acetate pH 7.75, 0.5 mM EDTA, 0.5 mg BSA, 0.2 mg polyvinylpyrrolidone (M_r 360.000), 10 μg D-luciferin (Biothema, Dalaro, Sweden), 5 mM magnesium acetate and 10 attomole to 0.01 attomole purified pyrophosphate or ATP. The addition of very low amounts of pyrophosphate or ATP (or analogs) was found to be important to decrease background light emission from the reaction mixture. Although the precise mechanism is unknown, BRC performed without adding small amounts of ATP or PPi consistently exhibited background luminescence that precluded accurate measurement of template nucleic acids present in amounts of about a femtomole or lower. Inorganic pyrophosphate present in the cDNA sample as a result of polymerase mediated dNTP incorporation was converted to ATP by sulfurylase. The ATP was used to generate light in a luciferin/luciferase reaction.

[0117] The generated light intensity over a time interval may be deconvoluted using the statistical analysis methods disclosed herein to determine the sequence of the template nucleic acid.

Synthesis and Purification of Sequence Specific Oligonucleotide Primers

[0118] The following oligonucleotides were synthesized and HPLC purified by MWG Biotech (High Points, NC).

B-MBPup

Biotin-5'-CGGCGATAAAGGCTATAACGG-3' (SEQ ID NO:2)

MBPup

5'-CGGCGATAAAGGCTATAACGG-3' (SEQ ID NO:3)

B-MBPR1

Biotin-5'-CTGGAACGCTTTGTCCGGGG-3' (SEQ ID NO:4)

MBPR1

5'-CTGGAACGCTTTGTCCGGGG-3' (SEQ ID NO:5)

oligo-loop

5'TTTTTTTTTTTTTTTTTTTTGCTGGAATTCGTCAGACTGGCCGTCGTTT

TACAACGGAACGGCAGCAAAATGTTGC-3' (SEQ ID NO:6)

Template Preparation

[0119] Biotinylated PCR products were prepared from bacterial extracts containing pMAL vector (New England Biolabs, Beverly, MA) (Pourmand et al. 1998, Autoimmunity 28; 225-233) by standard techniques, using MBPup and biotinylated B-MBPR1 or MBPR1 and biotinylated B-MBPup as PCR primers. The PCR products were immobilized onto streptavidin-coated superparamagnetic beads (Dynabeads™ M280-Streptavidin, Dynal A.S., Oslo, Norway). Single-stranded DNA was obtained by incubating the immobilized PCR product in 0.10 M NaOH for 3 min to separate strands and then removing the supernatant.

Strand Extension

[0120] The immobilized single stranded PCR product was resuspended in annealing buffer (10 mM Tris-acetate pH 7.75, 2 mM Mg-acetate) and placed into wells of a microtiter plate. Five pmol of the BRC primers MBP-up (SEQ ID NO:3) or MBPR1 (SEQ ID NO:5) were added to the immobilized strand obtained from the PCR reaction (depending on what set of biotinylated PCR primers was used). Hybridization of the template and primers was performed by incubation at 95°C for 3 min, 72°C for 5 min and then cooling to room temperature. Extension occurred in the presence of 10 U exonuclease-deficient (exo-) Klenow DNA polymerase (New England Biolabs, Beverly, MA) and addition of all four deoxynucleoside triphosphates to the extension mixture (0.14 mM final concentration). As discussed above, α-thio dATP was substituted for dATP to prevent interference with the luciferase reaction.

[0121] In an exemplary embodiment, extension and real-time luminometric monitoring were performed at 25°C in a Xenogen instrument (Xenogen, Menlo Park, CA). A luminometric reaction mixture was added to the substrate with different concentrations of extended primed single-stranded DNA or self primed oligonucleotide. The luminometric assay mixture (40 μl) contained 0.4 μg luciferase (Promega, Madison, WI), 15 mU recombinant ATP sulfurylase (Sigma Chemicals, St. Louis, MO), 0.1 M Tris-acetate (pH 7.75), 0.5 mM EDTA, 5 mM Mg-acetate, 0.1% (w/v) bovine serum albumin, 1 mM dithiothreitol, 10 μM adenosine 5'-phosphosulfate (APS) (Biolog, Alexis Biochemicals, Carlsbad, CA), 0.4 mg polyvinylpyrrolidone/ml (molecular weight 360000) and 100 μg D-luciferin/ml (BioThema AB, Haninge, Sweden). Emitted light was detected in real-time (data not shown). Using the

modified protocol with 0.01 attomole to 10 attomole purified pyrophosphate or ATP added, the background light intensity is essentially zero. Even in the presence of random noise background that is of approximately the same order of magnitude as the actual signal, the pyrophosphate induced signal can still be detected as a shift in the baseline level of the light output (not shown).

Detection Devices

[0122] The number of the photons generated by BRC may be measured using any known type of photodetector. Common devices that may be used include photodiodes, photomultiplier tubes (PMTs), charge coupled devices (CCDs), and photo-resistive materials. Luciferase-catalyzed photon generation has a quantum yield (Q.E.) of approximately 0.88, with the wavelength maximum depending on the type of luciferase used. For various types of luciferase, that can be anyplace within the visible range of the spectrum. Exemplary embodiments used firefly luciferase, which has a maximum intensity at 562 nm.

[0123] The photosensitive device is typically either in direct proximity of the BRC reaction to directly receive incident photons, or relatively far from the buffer with a light coupling device (*e.g.* optical fiber or mirror system) capable of directing light from the sample to the detector. In an exemplary embodiment, a UDT-PIN-UV-50-9850-1 photodiode (Hamamatsu Corp., Hamamatsu, Japan) was used with a transimpedance amplifier with a gain of 10^8 volts/amp.

Example 5: SNP detection Using Total RNA Templates

[0124] SNPs have been detected by hybridization of total RNA incubated with gene specific or allele specific primers and/or probes (Higgins *et al*, *Biotechniques* 23:710-714, 1997; Newton *et al. Lancet* 2:1481-1483, 1989; Goergen *et al*, *J Med Virol* 43:97-102, 1994; Newton *et al*, *Nucleic Acids Res* 17:2503-2516, 1989). Using the methods disclosed herein, SNPs may be detected by BRC, using sequence specific extension primers designed to bind to the template with the 3' end of the primer located over the base of interest (SNP site). In preferred embodiments, the primer sequence is selected so that the end of the primer to which nucleotides will be attached is base-paired with the polymorphic site. In certain embodiments, where the SNP is located in a coding sequence, the primer may be allowed to hybridize to total RNA or polyadenylated mRNA. (Alternatively, to detect non-coding SNPs genomic DNA or PCR amplified genomic DNA may be used as the template.) The template/primer fragments are used as the substrate for a primer extension reaction (*e.g.*, Sokolov, *Nucleic Acids Res* 18:3671, 1989)

in the presence of reverse transcriptase. If a template sequence is present that is complementary to the sequence specific primer, extension occurs and pyrophosphate is generated. Extension products (PPi) are detected as disclosed above, allowing identification of the SNP in the template nucleic acid.

[0125] Typically SNPs exist in one of two alternative alleles. The allelic variant of the SNP may be identified by performing separate BRC reactions with primers specific for each of the SNP variants. In an alternative embodiment, the SNP allele may be identified using a gene specific primer that binds immediately upstream of the SNP site, allowing extension to occur in the presence of a single type of dXTP (or α-thio dATP). Extension will occur if the added dXTP is complementary to the SNP nucleotide.

Example 6: Isothermal or Thermal Amplification of Nucleic Acids and BRC

A variety of nucleic acid amplification methods can be used in combination with BRC. Genomic DNA, cDNA, mRNA or total cell RNA may be extracted, mixed with appropriate reagents for amplification and BRC reagents for detection and quantification in the same tube. In certain embodiments, the amplification step may be performed separately from BRC detection and quantification.

Polymerase Chain Reaction (PCR) Amplification

[0126] Genomic DNA is extracted, combined with dNTP, Mg, buffer, Taq Polymerase enzyme and sequence specific primers. The samples are cycled through 1-30 rounds of denaturation at 95°C, annealing at 40-70°C and extension at 72°C. An aliquot the PCR amplified sample is added to BRC assay mix and the amount of PPi generated quantified as a measure of the number of starting copies of sequence specific DNA present in the sample. Alternatively the PCR step can be combined with the BRC assay in one tube using a thermostable luciferase enzyme and Sulfurylase enzyme. In this method there is a coupling of amplification and detection/quantification of the target sequence. The number of PPi released in solution as a result of amplification is directly proportional to the length of the target sequence, and can be used to quantify the number of starting target nucleic acid in solution.

Results

[0127] Genomic DNA was amplified with primers specific to Maltose Binding Protein (MBP). An aliquot of the PCR product was diluted serially and assayed using the BRC method.

Images were taken with standard CCD sensor with 1 sec integration time (not shown). Alternatively a luminometer was used with 10 sec integration time (not shown).

[0128] BRC was used with a complex genomic background with and without amplification steps. A bacterial colony containing the Rho 52 gene in a plasmid was grown on an agar plate. A colony with less than 100,000 bacteria was isolated and placed into 4 tubes containing buffer. The tubes were heated to 95°C for 5 minutes and then the master mix containing Taq Polymerase, dNTPs, primers specific to Rho 52 and Mg was added. Tube 1 was heated to 95°C for 1 min, 55°C for 1 min and 72°C for 1 minute for one amplification cycle. Tubes 2, 3 and 4 were amplified using similar temperatures but for 10 cycles, 20 cycles, and 30 cycles, respectively. An aliquot of each was added to the BRC assay for PPi measurement. We were able to detect and quantitate target DNA in each tube (not shown). Reference samples had all reagents and biological substances except primers.

[0129] Other potential isothermal applications that may be combined with BRC include Ribo-SPIA (Nugen Technologies), NASBA, RCA (Amersham), Ebervine (Ambion), Invader (Third Wave Techonologies) as well as cleavage based assays.

Example 7: Pathogen Typing by BRC

[0130] In certain embodiments of the invention, BRC can be used to identify, type and/or quantify template pathogens in a sample. Total RNA or genomic DNA of the pathogenic organism may be incubated with pathogen specific primers. In some embodiments, a single primer may be specific for one type of pathogen, or may be specific for a family of pathogens. Alternatively, multiple primers specific for different sub-types of a family of pathogens may be used. After hybridization in a suitable buffer, primer extension occurs with either reverse transcriptase or DNA polymerase, as disclosed above. The presence of a template pathogen type, or a member of a family of pathogens, is detected by luminescence using BRC. The pathogen titer (number of pathogenic organisms) in the sample may be determined by photon integration over a time interval.

Example 8: Pathogen Typing by Rolling Circle

[0131] In various embodiments, BRC may be performed using a rolling circle replication process. In this case, a circular primer sequence is allowed to hybridize with either total RNA or genomic DNA, for example of a pathogen. (Banér *et al*, *Nucleic Acids Research*, 26:5073-5078,

1998). As discussed above, the primer may be specific for a single type of pathogen, or may react with a family of pathogenic organisms. Alternatively, multiple circular primers specific for different members of a family of pathogens may be used. After hybridization, an exonuclease is added to the solution. The exonuclease digests single-stranded RNA or DNA, leaving intact double stranded RNA or DNA. The double stranded nucleic acid acts as the substrate in a primer extension reaction as discussed above, using reverse transcriptase or DNA polymerase. Formation of PPi is monitored by BRC.

Example 9: Protein-Protein Interaction

[0132] In some embodiments, BRC may be used to detect and/or quantify protein-protein binding. A set of putative template proteins may be immobilized onto a surface, such as a nitrocellulose or nylon membrane or microtiter plate. A protein or peptide that binds to the template protein may be tagged with a short oligonucleotide, for example using a bifunctional cross-linking reagent. The oligonucleotide-tagged protein or peptide may be incubated with the putative template proteins under conditions allowing binding to occur. The remaining unbound proteins may be washed away and the presence of bound oligonucleotide detected by rolling circle reaction as discussed above (Banér et al., 1998), using circular oligonucleotide primers that are complementary to the short oligonucleotide tag. BRC may be used to detect and/or quantify the number of bound template proteins. The skilled artisan will realize that the disclosed method is not limited to protein-protein interactions, but may be applied to any binding pair interaction where one member of the pair may be tagged with a short oligonucleotide. The method may also be applied to arrays of putative template proteins, for example where *in vitro* translation has been used to create an array of candidate binding proteins from mRNAs.

Example 10: Gene Expression Profiling by Using Total RNA or cDNA

[0133] Total RNA or cDNA may be incubated with one or more gene specific primers or general primers. Bound primer/template pairs are extended by reverse transcription or DNA polymerization. Formation of pyrophosphate is detected by BRC, as discussed above. In certain embodiments, a primer is used that is designed to bind specifically to a single gene product (mRNA species), allowing determination of expression for an individual gene. In other embodiments, non-specific primers, such as oligo(dT) and/or random primers may be used. In this case, the mRNA species present in a sample may be first separated, for example by

hybridization to a DNA microarray containing complementary sequences for a large number of gene products. Hybridization may be followed by non-specific primer binding, extension and BRC reaction. Alternatively, the oligonucleotides of the array may themselves be used as primers, allowing extension and light emission to occur. In such embodiments, the PPi reaction product may preferably be localized so that light emission is limited to the immediate location of a hybridized template nucleic acid. Many such localization techniques are known in the art, for example using microtiter plates wherein each well contains a probe for an individual gene expression product, or using a commercial apparatus such as a Nanochip® Workstation (Nanogen, San Diego, CA).

Example 11: Real time PCR

[0134] The quantification of amplified template in a polymerase chain reaction (PCR) is achieved by incorporation of dNTP. As a result of dNTP incorporation PPi is released. An aliquot of synthesized DNA from each PCR cycle is added to a reaction mixture containing luciferase as disclosed above and thereby one can evaluate/estimate the mass of the molecules for each cycle from the generated light.

* * *

[0135] All of the COMPOSITIONS, METHODS and APPARATUS disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the COMPOSITIONS, METHODS and APPARATUS and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

REFERENCES

[0136] The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

[0137] Akeson, M., D. Branton, J.J. Kasianowicz, E. Brandin, and D.W. Deamer. 1999. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. Biophys. J. 77: 3227-3233.

[0138] Anastassiou D., "Genomic Signal Processing", *IEEE Signal Processing Magazine*, vol. 18, no.4, pp. 8-20, (2001)

[0139] Beynon, J., MacRae, I., Huston, S., Nelsom, D., Segel, I., Fisher, A., *Biochemistry*, 40, 14509-14517 (2001)

[0140] Brovko, L., Gandel'man, O., Polenova, T., Ugarova, N., Biochem. (Moscow), 59, 195-201 (1994)

[0141] Haykin S., *Adaptive Filtering*, Prentice-Hall, Englewood-Cliffs, NJ, (2002)

[0142] Kailath, T., Sayed, A.H., Hassibi, B., *Linear Estimation,* Prentice-Hall, Englewood-Cliffs, NJ, (2000)

[0143] Kasianowicz, J.J., E. Brandin, D. Branton, and D.W. Deamer. 1996. Characterization of individual polynucleotide molecules using a membrane channel. Proceedings of the National Academy of Sciences U.S.A. 93: 13770-13773.

[0144] Meller, A., L. Nivon, and D. Branton. 2001. Voltage-driven DNA translocations through a nanopore. Phys. Rev. Lett. 86: 3435-3438.

[0145] Meller, A., L. Nivon, E. Brandin, J. Golovchenko, and D. Branton. 2000. Rapid nanopore discrimination between single polynucleotide molecules. Proc. Natl. Acad. Sci. USA 97: 1079-1084.

[0146] Nyren, P. *Anal. Biochem.*, 167, 235-238 (1987)

[0147] Nyren, P., Lundin, A., *Anal. Biochem.*, 151, 504-509 (1985)

[0148] Nyren, P., Karamohamed, S., Ronaghi, M., *Anal. Biochem.*, 244, 367-373 (1997)

[0149] Ronaghi, M *et al. Anal. Biochem*, 242, 84-89 (1996)

[0150] Ronaghi, M, Uhlen, M., Nyren, P., *Science*, 281, 363-365 (1998)

[0151] Ronesto, F., Martin, R., Borrel, J., Nelson, D., Segel, I., *Arch. Biochem. Biophy.* 290, 66-78 (1994)

WHAT IS CLAIMED IS:

1. A method of nucleic acid sequencing comprising:

   a) replicating multiple copies of a nucleic acid template molecule;

   b) obtaining a complex signal from the nucleic acid replication;

   c) deconvoluting the signal by statistical signal processing; and

   d) determining the sequence of the template nucleic acid.

3. The method of claim 1, wherein the complex signal is obtained by BRC (bioluminescence regenerative cycle) detection.

4. The method of claim 1, wherein the complex signal is obtained by charge perturbation signature analysis.

5. The method of claim 1, wherein the complex signal is obtained by electrical detection with nanopores.

6. The method of claim 1, wherein the signal deconvolution comprises performing a Wiener solution analysis.

7. The method of claim 6, wherein the complex signal reflects a uniformly distributed time delay.

8. The method of claim 6, wherein the complex signal reflects a non-uniform delay distribution.

9. The method of claim 6, further comprising interatively estimating the N_j .

10. The method of claim 1, wherein the signal is deconvoluted using nonlinear techniques.

ABSTRACT

[0152] The present invention concerns methods for bio-molecular sequence recovery using statistical signal processing. In particular embodiments, the bio-molecules are nucleic acids and nucleic acid sequences are obtained. In various embodiments, the statistical signal processing methods may comprise Wiener solution analysis, exploiting large numbers of template molecules, uniform or non-uniform delay distribution, and linear or non-linear statistical methods. The signals may be detected by BRC (bioluminescence regenerative cycle), charge perturbation signature analysis, electrical detection with nanopores, or any other known method of signal generation for nucleic acid sequencing.

Figure 1: Waveforms for each of the ATGC bases.

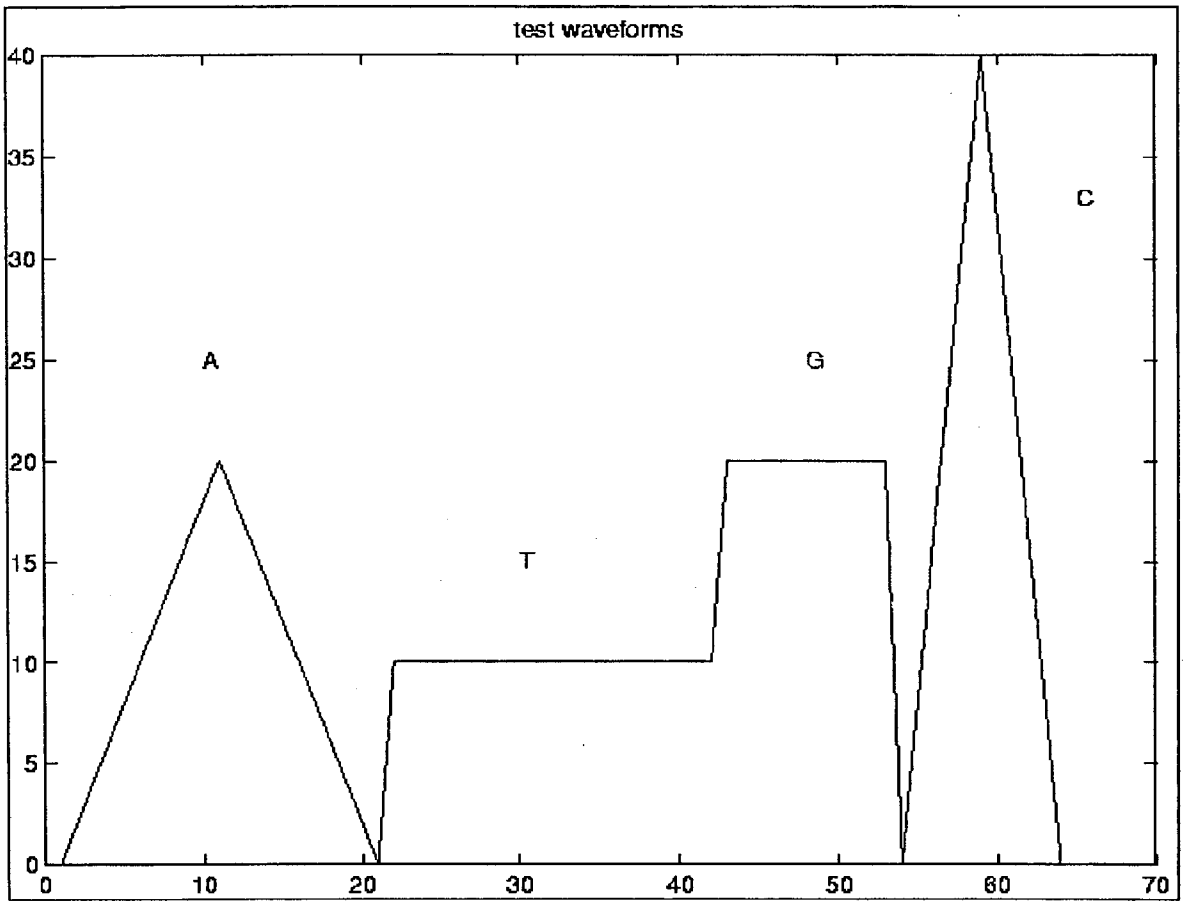


Figure 2: Waveform for the reaction of the TCTAGCTCAG sequence.

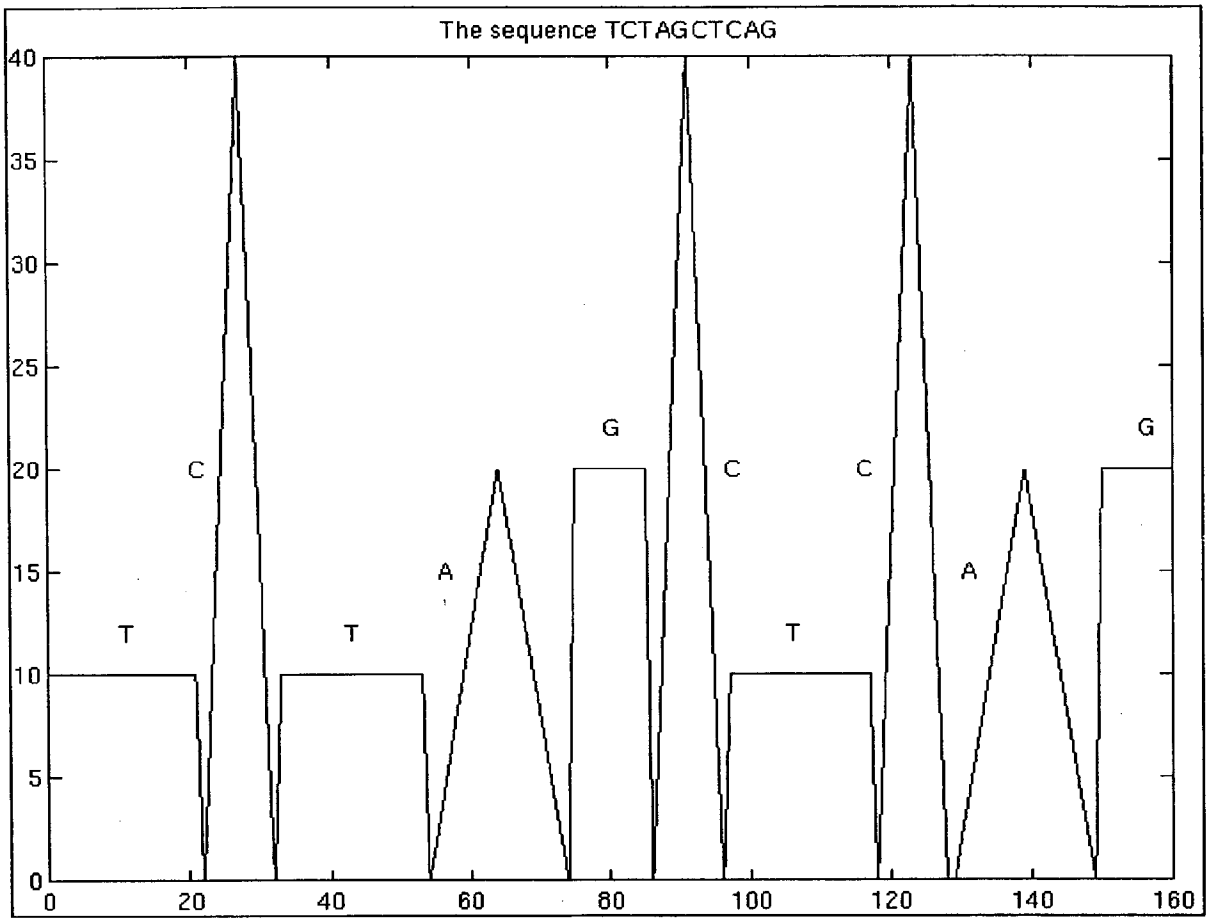


Figure 3: Noise-corrupted aggregate waveform obtained from the uniformly asynchronous reaction of 10^5 DNA molecules with TCTAGCTCAG sequence.

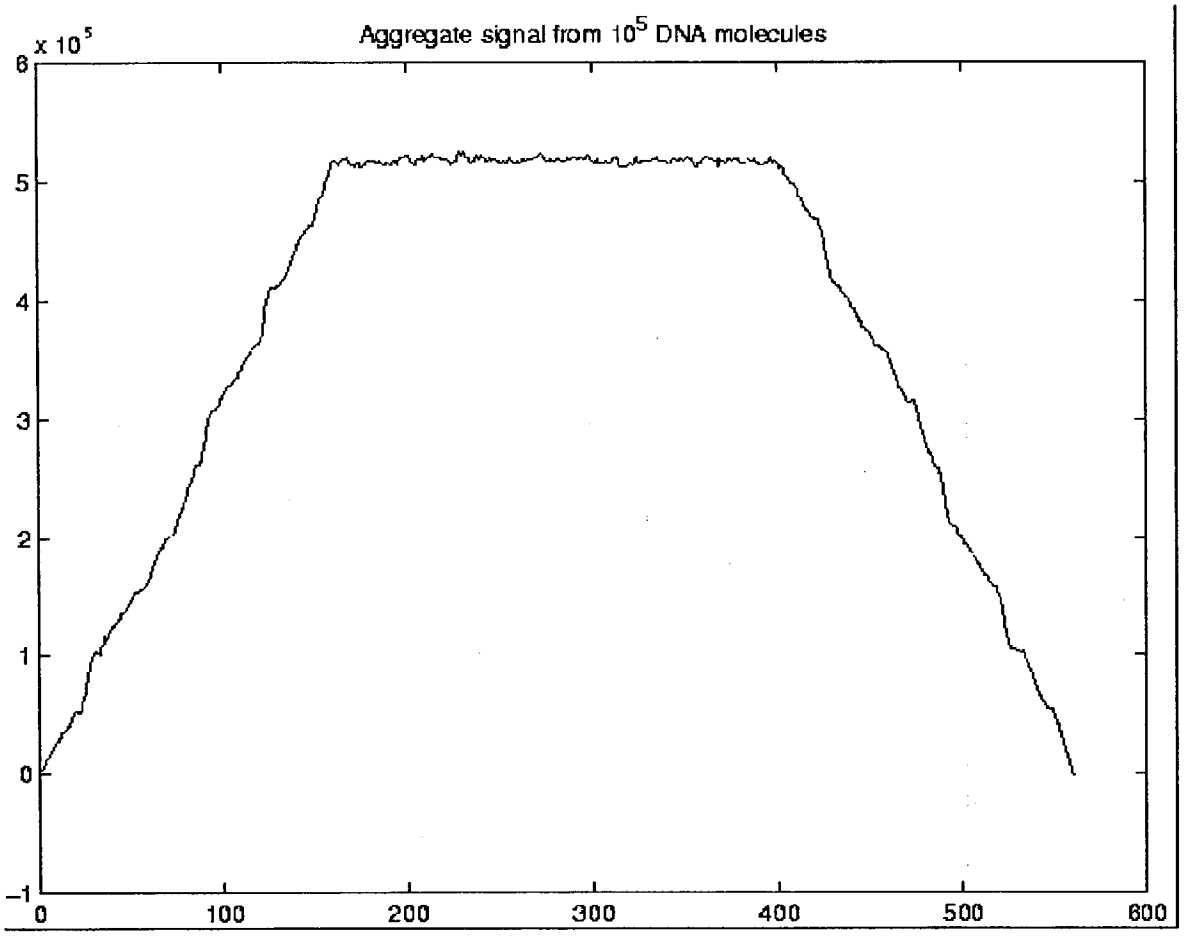


Figure 4: Reconstructed waveform using the Wiener solution ($SNR_{perfect} = 40db$).

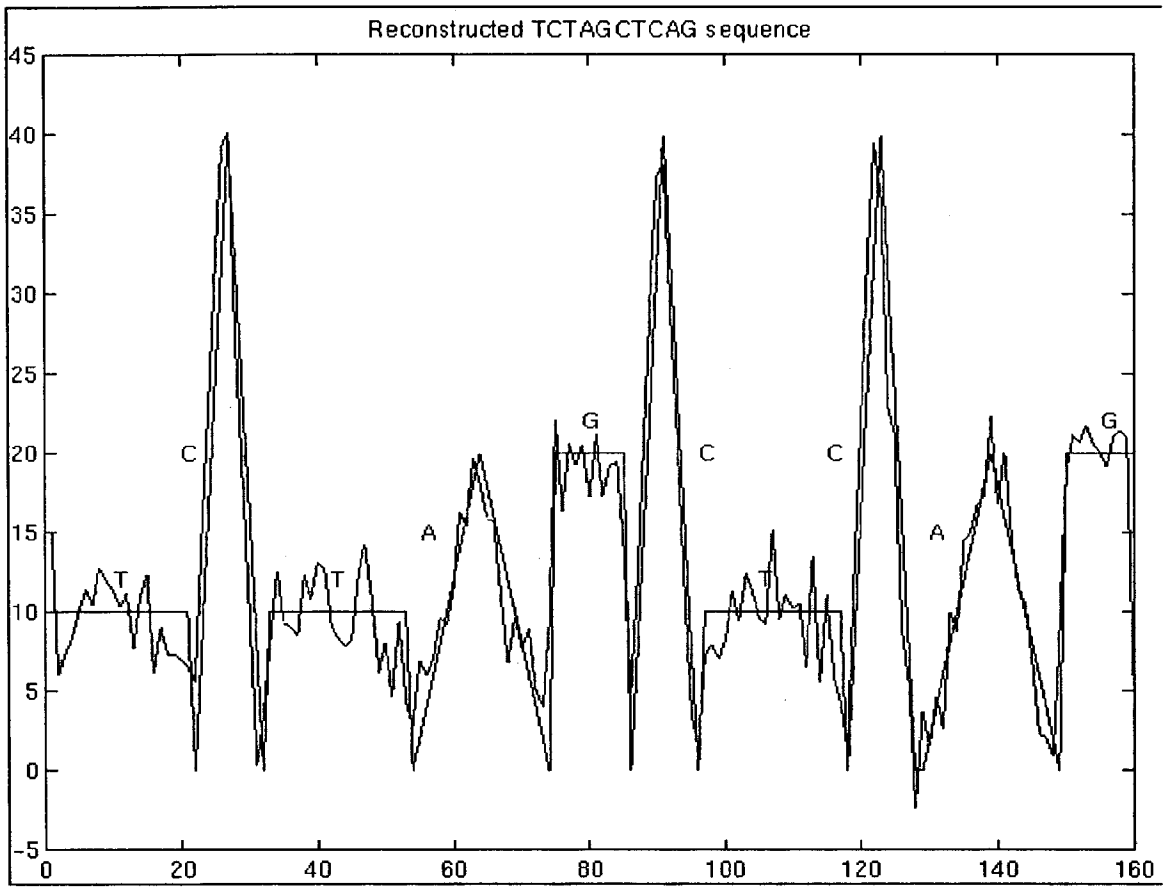


Figure 5: Reconstructed waveform using the Wiener solution ($SNR_{perfect} = 35db$).

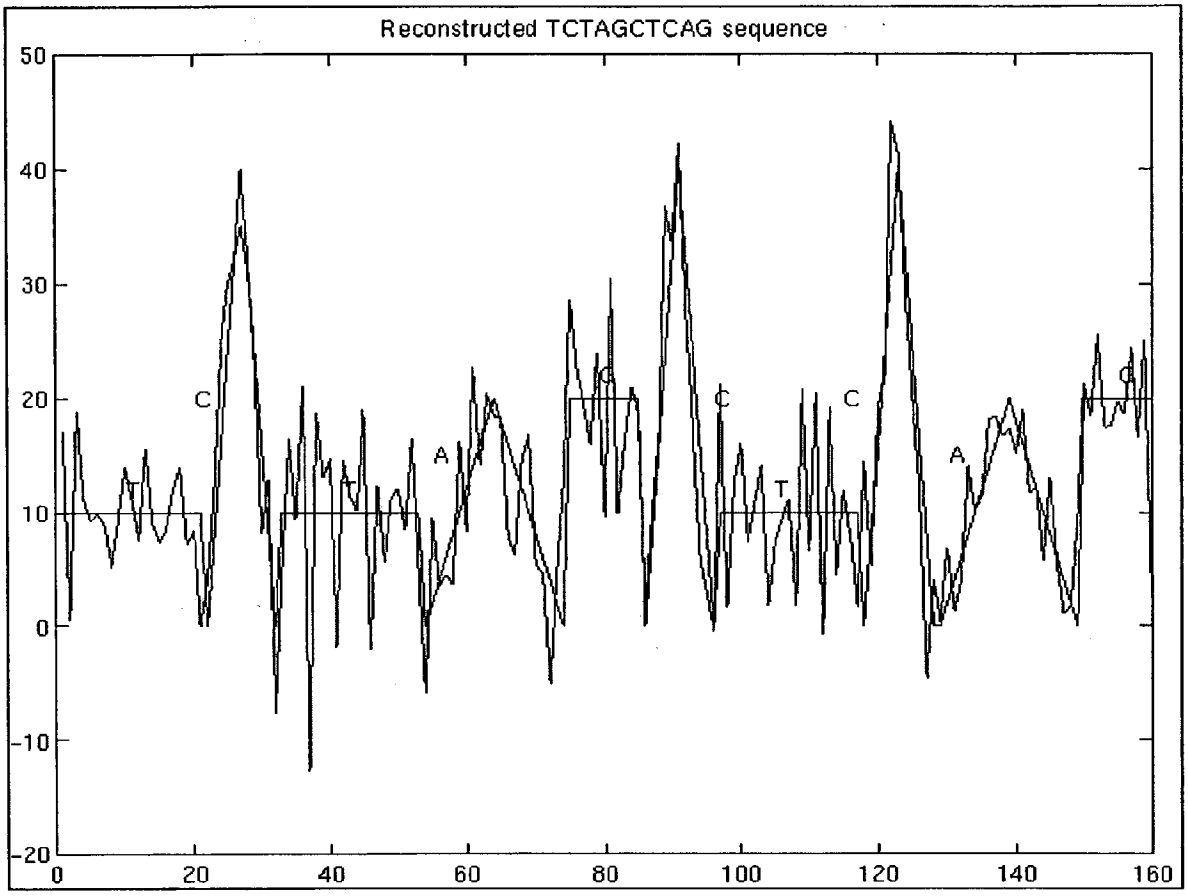


Figure 6: Reconstructed waveform using the Wiener solution ($SNR_{perfect} = 30db$).

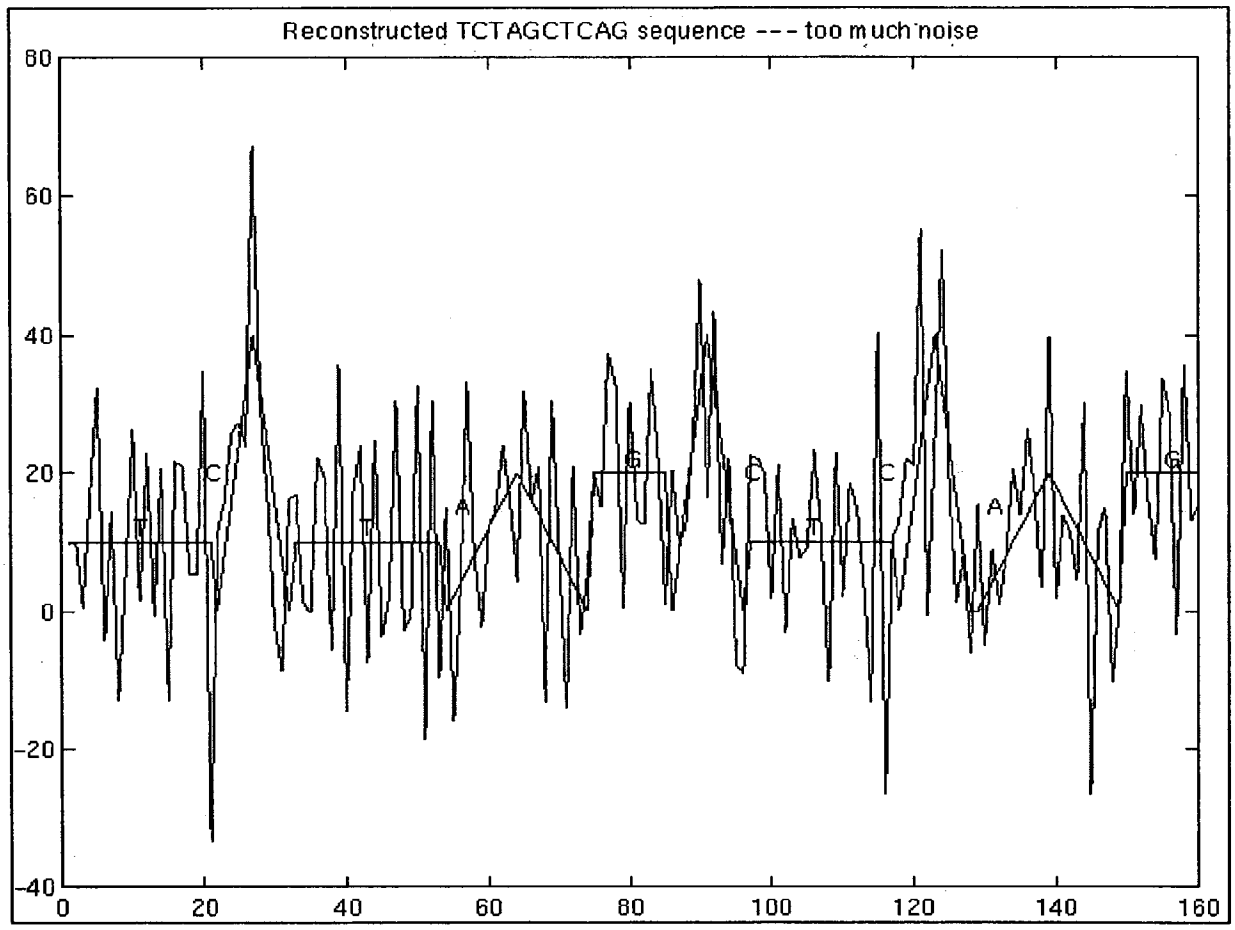


Figure 7: Reconstructed waveform using the Wiener solution ($SNR_{perfect} = 40db$ and $N = 10^6$).

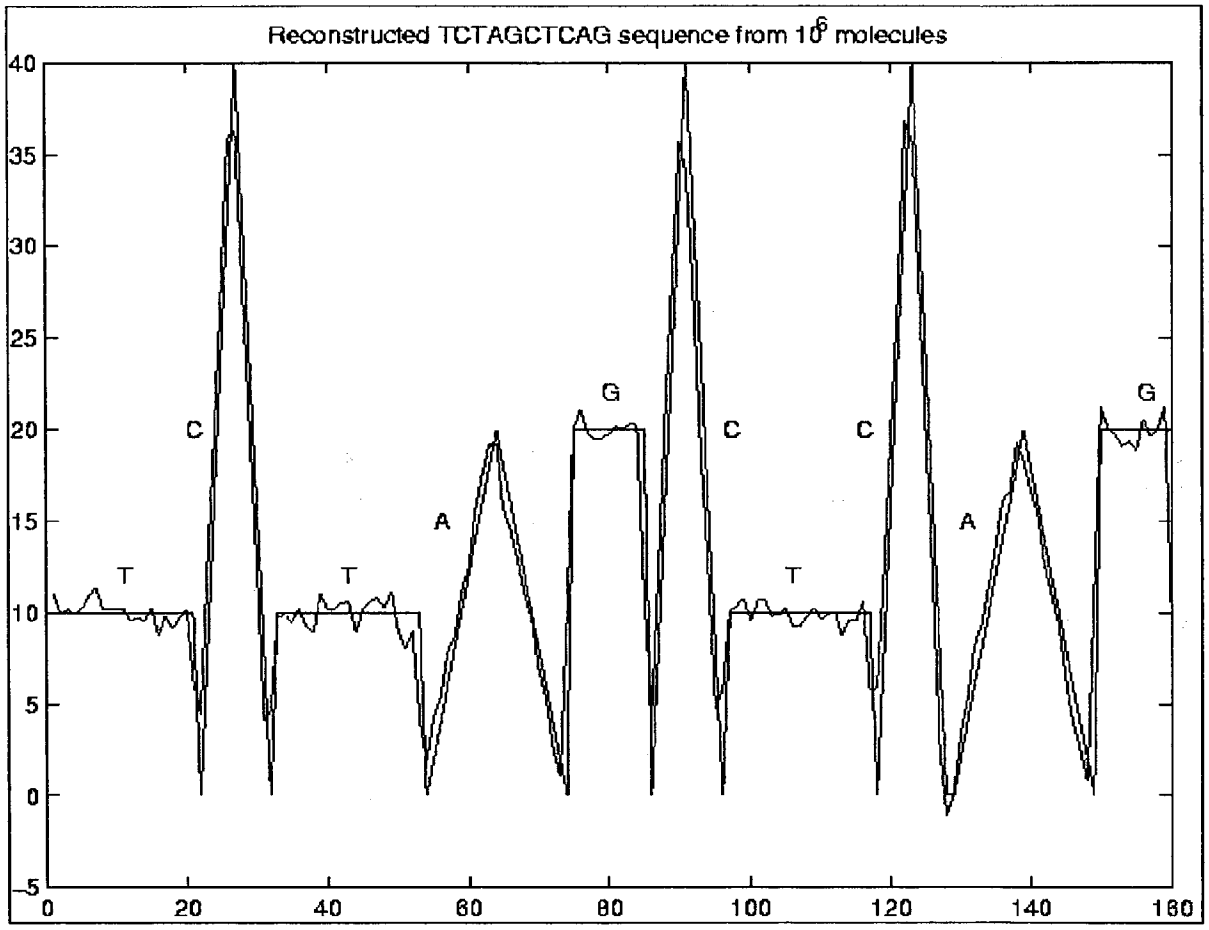


Figure 8: Noise-corrupted aggregate waveform of 10^5 DNA molecules with Gaussian delay distribution.

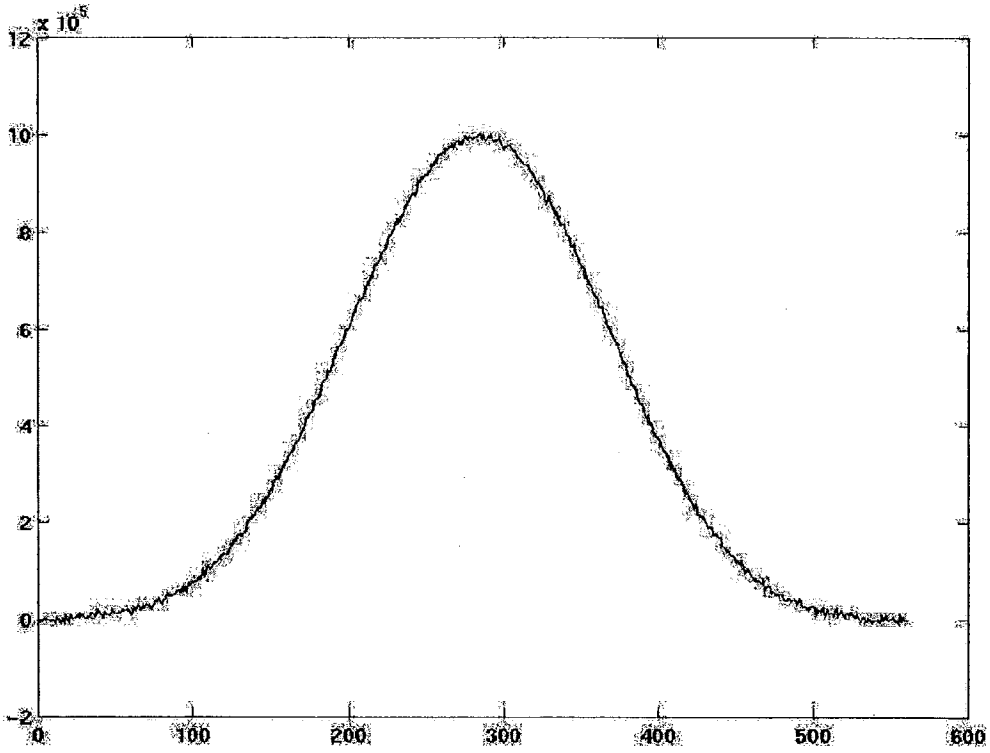


Figure 9: Reconstructed waveform using the Wiener solution when the delay distribution is Gaussian ($SNR_{perfect} = 40db$).

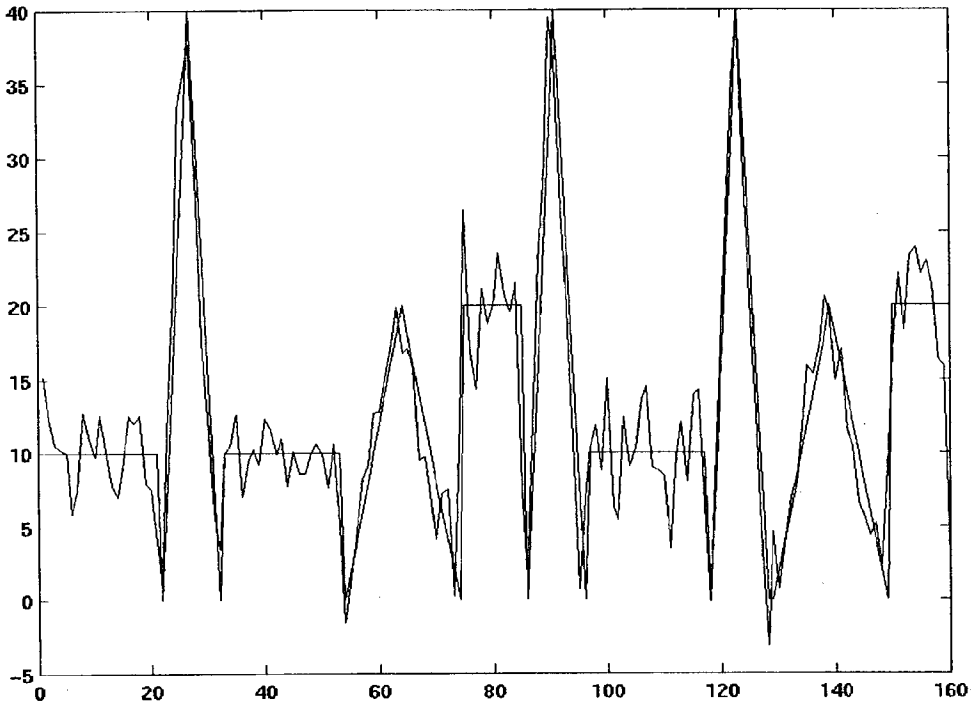


Figure10: A typical Nanopore structure and current modulation due to blocking effect of a DNA strand.

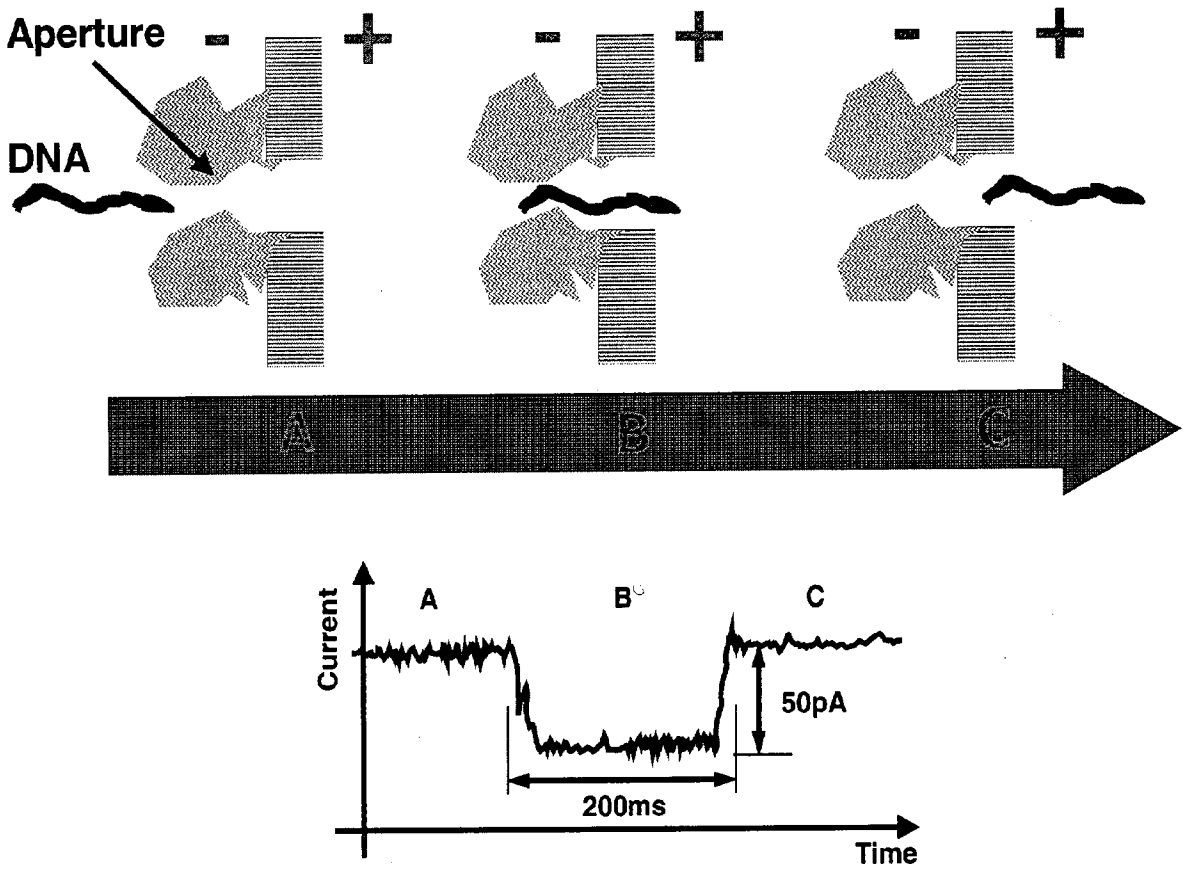


Figure11: A multi-aperture nanopore structure and current modulation due to blocking effect of multiple identical DNA strands.

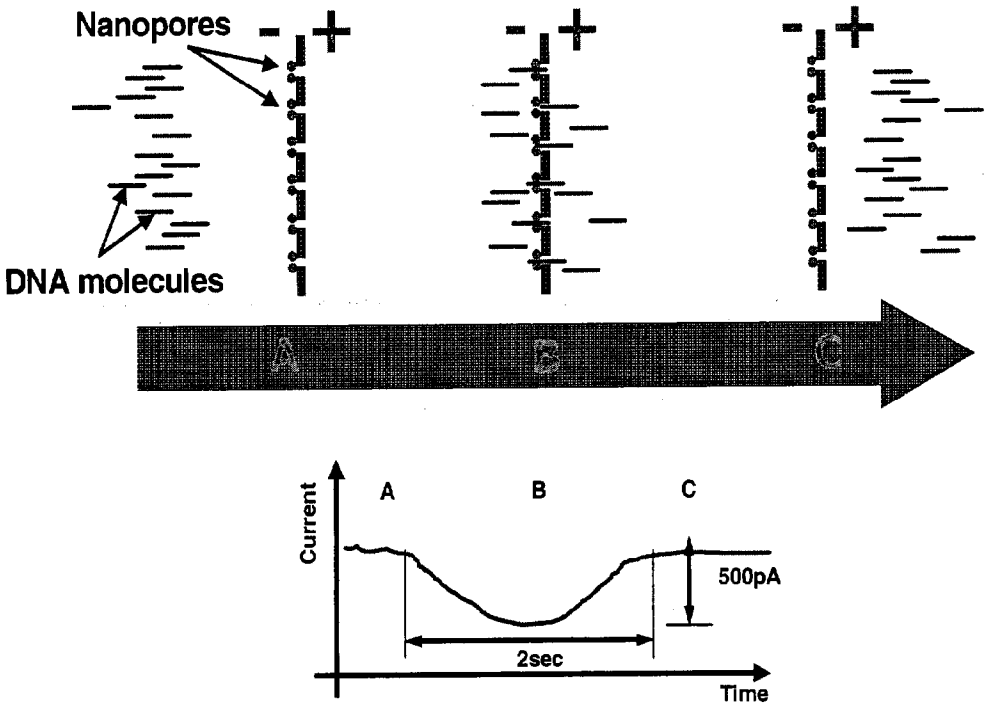
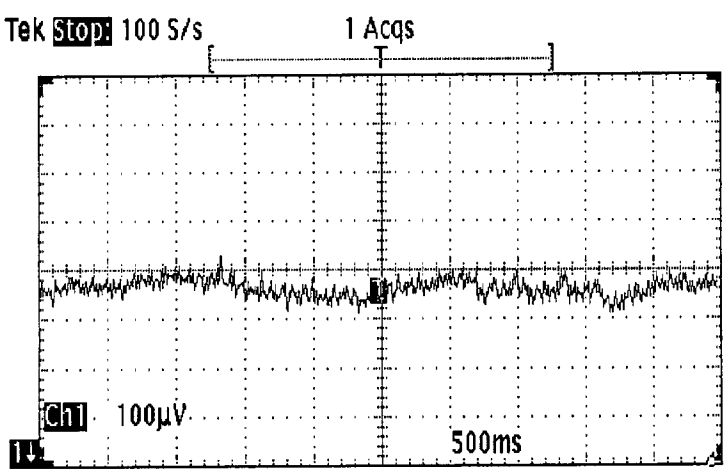


Figure12: (A) Background noise of the solution, (B) Output of the sensor during polymerization.



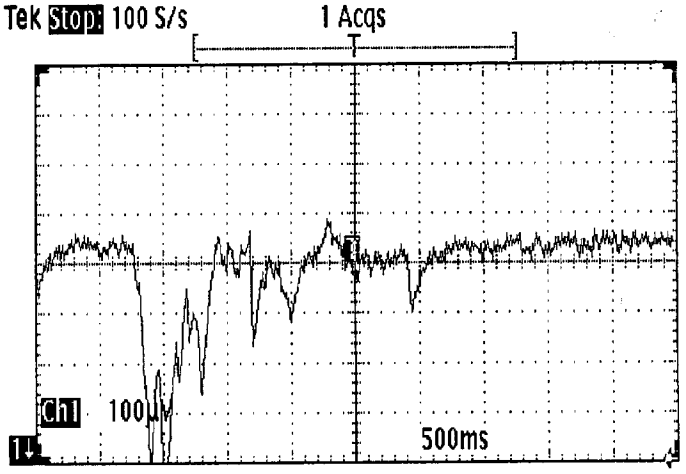
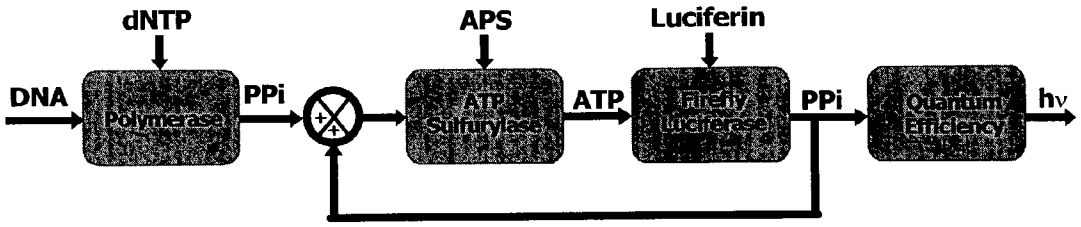


Figure13: Block diagram of bioluminescence regenerative cycle process.



PATENT

Attorney Docket: 005852.P006Z

APPLICATION FOR UNITED STATES LETTERS PATENT

for

MODULATION OF LUCIFERASE ACTIVITY IN THE BIOLUMINESCENCE REGENERATIVE CYCLE (BRC) ASSAY WITH A PEPTIDE LINKER

by

Nader Pourmand

Arjang Hassibi

and

Siavash Ghazvini

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 Wilshire Boulevard, Seventh Floor
Los Angeles, California 90025-1030

Telephone: (303) 740-1980
Facsimile: (303) 740-6962

EXPRESS MAIL CERTIFICATE OF MAILING
"Express Mail" Mailing No.: EL 880593797 US

I hereby certify that I am causing the above-referenced correspondence to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated below and that this paper or fee has been addressed to Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit: December 20, 2002

Name of Person Mailing Correspondence: Lori E. True

Lori E. True
Signature

December 20, 2002
Date

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to the field of protein and/or peptide detection and/or quantification. More particularly, the present invention concerns novel approaches to detection and/or quantification of peptides, proteins, enzymes and/or proteases using a bioluminescence regenerative cycle (BRC) assay, with modulation of luciferase and/or ATP Sulfurylase activity by one or more peptide linkers.

Description of Related Art

[0002] Methods of precise and highly sensitive detection and/or quantification of proteins and/or peptides are of use for a variety of medical, forensic, epidemiological, public health, biological warfare and other applications. A variety of techniques would benefit from the availability of precise and sensitive methods for protein and/or peptide detection and/or quantification.

[0003] Present methods for detection, identification and/or quantification of individual proteins or peptides are primarily based on binding of selective or specific antibodies to the protein of interest. Well-known antibody-based methods include ELISA, sandwich ELISA, Western blot, slot blot, dot blot, microtiter plate immunoassay, immunohistochemical analysis and related techniques. Protein analysis may also be performed by non-antibody based assays. Where a specific enzyme is to be detected, a chromogenic or other enzyme activity assay may be designed and used. More general assays may be based on SDS-polyacrylamide gel electrophoresis, mass-spectrometry, nuclear magnetic resonance analysis, ligand binding assays or other alternatives.

[0004] Many of these present methods suffer from lack of specificity and/or sensitivity. Antibody-based assays may show non-specific cross-reactivity, with antibodies binding to proteins other than the protein of interest. This is particularly true where the protein of interest is present in low abundance in samples, compared to other cross-reactive proteins. Enzyme based assays also can exhibit cross-reactivity between enzymes that catalyze similar reactions. Interpretation of some types of assay, such as gel electrophoresis or mass spectrometry, may be difficult if the protein of interest forms a small fraction of the total amount of protein in a

sample. A need exists in the art for a sensitive and specific method of identifying, detecting and/or quantifying proteins or peptides in a sample.

[0005] Pyrophosphate based detection systems have been reported, primarily for use in nucleic acid analysis (*e.g.*, Nyren and Lundin, *Anal. Biochem.* 151:504-509, 1985; U.S. Patent Nos. 4,971,903; 6,210,891; 6,258,568; 6,274,320, each incorporated herein by reference). The systems use a coupled reaction wherein pyrophosphate is generated by an enzyme-catalyzed process, such as nucleic acid polymerization. The pyrophosphate is used to produce ATP, in an ATP sulfurylase catalyzed reaction with adenosine 5'-phosphosulphate (APS). The ATP in turn is used for the production of light in a luciferin-luciferase coupled reaction. The present invention provides a novel method of pyrophosphate-based detection for use in detection, identification and/or quantification of proteins and/or peptides.

SUMMARY OF THE INVENTION

[0006] The present invention fulfills an unresolved need in the art by providing methods for accurately detecting and/or quantifying target analytes, such as proteins and/or peptides, in samples. In preferred embodiments, the number of target proteins or peptides in a sample may be accurately determined over a wide concentration range. The disclosed methods provide increased sensitivity and accuracy of target molecule quantification compared to prior art methods. The disclosed methods involve, in part, a "bioluminescence regenerative cycle" or BRC. In preferred embodiments, the activity of the BRC process is initially inhibited by the presence of a selected peptide covalently or non-covalently attached to one or more of the BRC enzymes, such as luciferase or ATP sulfurylase. Removal of the inhibitory peptide by a protein or peptide of interest, present in a sample to be analyzed, initiates the light emitting BRC reactions. In some embodiments the inhibitory peptide may be removed by a target protease in the sample. In other cases, another type of target protein or peptide may act to restore activity of the inhibited BRC enzyme.

[0007] In certain embodiments of the invention, the methods may comprise obtaining at least one sample suspected of containing one or more target proteins and/or peptides, initiating pyrophosphate generation by activating one or more BRC enzymes, producing light by a bioluminescence regenerative cycle, accumulating the total number of photons produced over

different time intervals, comparing the photon accumulation with the background photon emission and determining the number of target proteins and/or peptides in the sample

[0008] In other embodiments of the invention, a target protease may be covalently or non-covalently attached to another molecule to be quantified, such as another protein, peptide or other ligand. The protein, peptide or ligand may be indirectly quantified, by detecting the attached protease. Such protease tagged ligands may be used, for example, to quantify protein-protein binding interactions or any other type of known ligand-receptor binding interaction. The methods are not limited by the type of target protease used, including but not limited to a serine protease, a cysteine protease, an aspartic protease, a metallo-protease, a cathepsin, a collagenase, an elastase, kallikrein, plasmin, renin, streptokinase, subtilisin, thermolysin, thrombin, urokinase, HIV protease, trypsin, chymotrypsin, pepsin, gastrin, calcium-dependent proteases, magnesium-dependent proteases, proteinase K, papain, bromelain, or any other protease known in the art. The specificities of various proteases for different target peptide sequences are well known in the art. In certain embodiments, the presence or amount of a specific protease in a sample may be diagnostic for a disease state, such as cancer or hemophilia. In other embodiments, the presence of a bacterial or viral encoded protease in a sample, such as HIV protease or streptokinase, may be diagnostic for the presence of an infection with a pathogenic organism.

[0009] In some embodiments, the pyrophosphate producing reaction is allowed to proceed to completion before BRC analysis. Once the reaction is complete, the pyrophosphate is reacted with APS in the presence of ATP sulfurylase to produce ATP and sulphate. The ATP is reacted with oxygen and luciferin in the presence of luciferase to yield oxyluciferin, AMP and pyrophosphate. For each molecule of pyrophosphate that is cycled through BRC, a photon of light is emitted and one molecule of pyrophosphate is regenerated. Because of the relative kinetic rates of luciferase and ATP sulfurylase, a steady state is reached in which the concentrations of ATP and pyrophosphate and the level of photon output remain relatively constant over an extended period of time. The number of photons may be counted over a time interval to determine the number of target proteins and/or peptides in the sample. The very high sensitivity of BRC is related in part to the integration of light output over time, in contrast to other methods that measure light output at a single time point or at a small number of fixed time points. The ability to vary the length of time over which photon integration occurs also

contributes to the very high dynamic range for protein or peptide quantification. The detection noise is also significantly reduced by increasing the length of integration.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0011] **FIG. 1** illustrates an exemplary method for modulation of luciferase activity in the BRC assay with a peptide linker.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Definitions

[0012] Terms that are not otherwise defined herein are used in accordance with their plain and ordinary meaning.

[0013] As used herein, "a" or "an" may mean one or more than one of an item.

[0014] As used herein, "luminescence" refers to the emission of light that does not derive energy from the temperature of the emitting body (i.e., emission of light other than incandescent light). "Luminescence" includes, but is not limited to, fluorescence, phosphorescence, thermoluminescence, chemiluminescence, electroluminescence and bioluminescence. "Luminescent" refers to an object that exhibits luminescence. In preferred embodiments, the light is in the visible spectrum. However, the present invention is not limited to visible light, but includes electromagnetic radiation of any frequency.

[0015] As used herein, the terms "analyte" and "target" mean any compound, molecule or aggregate of interest for detection. Non-limiting examples of targets include a peptide, polypeptide, protein, carbohydrate, polysaccharide, glycoprotein, lipid, hormone, growth factor, cytokine, receptor, antigen, allergen, antibody, substrate, metabolite, cofactor, inhibitor, drug, pharmaceutical, nutrient, toxin, poison, explosive, pesticide, chemical warfare agent, biowarfare agent, biohazardous agent, infectious agent, prion, radioisotope, vitamin, heterocyclic aromatic compound, carcinogen, mutagen, narcotic, amphetamine, barbiturate, hallucinogen, waste

product, contaminant, heavy metal or any other molecule or atom, without limitation as to size. "Targets" are not limited to single molecules or atoms, but may also comprise complex aggregates, such as a virus, bacterium, *Salmonella, Streptococcus, Legionella, E. coli, Giardia, Cryptosporidium, Rickettsia,* spore, mold, yeast, algae, amoebae, dinoflagellate, unicellular organism, pathogen or cell. In certain embodiments, cells exhibiting a particular characteristic or disease state, such as a cancer cell, may be targets. Virtually any chemical or biological compound, molecule or aggregate could be a target.

BRC Detection Method

[0016] Various embodiments of the invention concern novel methods for quantifying proteins and/or peptides without labeling of any target, capture or probe molecules. Such label free methods are advantageous with respect to sensitivity, expense and ease of use. The BRC methods involve the luminescent detection of pyrophosphate (PPi) molecules released from an enzyme-catalyzed reaction, such as RNA and/or DNA polymerization. As part of the technique, a bioluminescence regenerative cycle (BRC) is triggered by the release of inorganic pyrophosphate (PPi) from polymerization of a protein and/or peptide.

[0017] The BRC regenerative cycle can be utilized with any reaction that generates pyrophosphate (PPi), such as nucleic acid polymerization. The PPi generated reacts with APS, catalyzed by ATP-sulfurylase enzyme, which results in the production of ATP and inorganic sulphate. In another reaction, luciferin and luciferase consume ATP as an energy source to generate light, AMP and oxyluciferin and to regenerate PPi. Thus, after each BRC cycle, a quantum of light is generated for each molecule of PPi in solution, while the net concentration of ATP in solution remains relatively stable and is proportional to the initial concentration of PPi. In the course of the reactions, APS and luciferin are consumed and AMP and oxyluciferin are generated, while ATP sulfurylase and luciferase remain constant. The invention is not limited as to the type of luciferase used. Although certain disclosed embodiments utilized firefly luciferase, any luciferase known in the art may be used in the disclosed methods.

[0018] As a result of the BRC process, the photon emission rate remains steady and is a monotonic function of the amount of PPi in the initial mixture. For very low concentrations of PPi (10^{-8} M or less), the total number of photons generated in a fixed time interval is proportional to the number of PPi molecules. Where PPi is generated by the polymerase catalyzed replication

of a target protein and/or peptide, the number of photons generated in a fixed time interval is proportional to the quantity of the target protein and/or peptide present in the sample.

[0019] The basic concept of enzymatic light generation from PPi molecules was introduced almost two decades ago (Nyren and Lundin, *Anal. Biochem.* 151:504-509, 1985; Nyren, *Anal. Biochem.* 167:235-238, 1987). Pyrophosphate based luminescence has been used for DNA sequencing (Ronaghi *et al.*, *Anal. Biochem.* 242:84-89, 1996) and SNP detection (Nyren *et al.*, *Anal. Biochem.* 244:367-373, 1997). The present methods provide additional procedures for accurately detecting, identifying and/or quantifying specific target proteins and/or peptides in the presence of contaminants and detector noise. The novel system and methods have an intrinsic controllable dynamic range up to seven orders of magnitude.

Proteins and Peptides

[0020] In different embodiments of the invention, protein and/or peptides to be detected, identified and/or quantified may be: [1] purified from natural sources; [2] expressed by *in vitro* translation of an mRNA species or by linked transcription/translation of a DNA species; and/or [3] expressed in a host cell that has been transformed with a gene or a complementary DNA (cDNA) species. These methods are not limiting and protein and/or peptides to be detected, identified and/or quantified may be prepared by any method known in the art.

Protein Purification

[0021] In certain embodiments of the invention, proteins and/or peptides to be detected, identified and/or quantified may be partially or fully purified from a variety of sources before analysis. Protein purification techniques are well known to those of skill in the art. These techniques typically involve an initial crude fractionation of cell or tissue homogenates and/or extracts into protein and non-protein fractions. Fractionation may utilize, for example, differential solubility in aqueous solutions, detergents and/or organic solvents, elimination of classes of contaminants such as proteins and/or peptides by enzymatic digestion, precipitation of protein and/or peptides with ammonium sulphate, polyethylene glycol, antibodies, heat denaturation and the like, followed by ultracentrifugation. Low molecular weight contaminants may be removed by dialysis, filtration and/or organic phase extraction.

[0022] Protein(s) of interest may be purified using chromatographic and/or electrophoretic techniques to achieve partial or complete purification. Methods suited to the purification of

protein and/or peptides and peptides include, but are not limited to, ion-exchange chromatography, gel exclusion chromatography, polyacrylamide gel electrophoresis, affinity chromatography, immunoaffinity chromatography, hydroxylapatite chromatography, hydrophobic interaction chromatography, reverse phase chromatography, isoelectric focusing, fast protein liquid chromatography (FPLC) and high pressure liquid chromatography (HPLC). These and other methods of protein purification are known in the art and are not limiting for the claimed subject matter. Any known method of protein purification may be used. There is no requirement that the protein must be in its most purified state. Methods exhibiting a lower degree of relative purification may, for example, have advantages in increased recovery of labeled protein.

[0023] Particular embodiments of the invention may rely on affinity chromatography for purification of protein and/or peptides. The method relies on an affinity between a protein and a molecule to which it can specifically bind. Chromatography material may be prepared by covalently attaching a protein-binding ligand, such as an antibody, antibody fragment, receptor protein, substrate, inhibitor, product or an analog of such ligands to an insoluble matrix, such as column chromatography beads or a nylon or other membrane. The matrix is then able to specifically adsorb the target protein from a solution. Elution occurs by changing the solvent conditions (*e.g.* pH, ionic strength, temperature, detergent concentration, *etc.*). One of the most common forms of affinity chromatography is immunoaffinity chromatography. Methods for generating antibodies against various types of protein and/or peptides for use in immunoaffinity chromatography are well known in the art.

[0024] In some embodiments of the invention, one or more protein and/or peptides of interest may be specifically labeled in order to facilitate purification. The protein of interest may be followed through a purification protocol by looking for the presence of labeled protein. In other embodiments of the invention, proteins may be post-translationally labeled using side chain specific and/or selective reagents as discussed below. Various methods for protein labeling are known in the art.

In Vitro Translation

[0025] Protein and/or peptides may be expressed using an *in vitro* translation system with mRNA templates. Complete kits for performing *in vitro* translation are available from

commercial sources, such as Ambion (Austin, TX), Promega (Madison, WI), Amersham Pharmacia Biotech (Piscataway, NJ), Invitrogen (Carlsbad, CA) and Novagen (Madison, WI). Such kits may utilize total RNA, purified polyadenylated mRNA, and/or purified individual mRNA species obtained from a cell, tissue or other sample. Methods of preparing different RNA fractions and/or individual mRNA species for use in *in vitro* translation are known. (*E.g.*, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; Ausubel et al., Current Protocols in Molecular Biology, Wiley and Sons, New York, NY, 1994).

[0026] Commonly used *in vitro* translation systems are based on rabbit reticulocyte lysates, wheat germ extracts and *E. coli* extracts. *In vitro* translation systems based on rabbit reticulocyte lysates are particularly robust and efficient for eukaryotic translation. The systems contain crude cell extracts including ribosomal subunits, transfer RNAs (tRNAs), aminoacyl-tRNA synthetases, initiation, elongation and termination factors and/or all other components required for translation. In certain embodiments of the invention, the natural amino acids present in such extracts may be supplemented with one or more different types of labeled amino acids. Depending on the embodiment, the label may be restricted to a single type of amino acid. Alternatively, a sample to be translated may be divided up into different sub-samples, each of which may be exposed to a different type of labeled amino acid. Other components of use in supplementing *in vitro* translation systems and methods of use of such systems are known in the art (*e.g.*, www.ambion.com/basics/translation/translation101.html).

[0027] In certain alternative embodiments of the invention, *in vitro* translation may be linked to transcription of genes to generate mRNAs. Such linked transcription/translation systems may use PCR® amplification products and/or DNA sequences inserted into standard expression vectors such as BACs (bacterial artificial chromosomes), YACs (yeast artificial chromosomes), cosmids, plasmids, phage and/or other known expression vectors. Linked transcription/translation systems are available from commercial sources (*e.g.*, Proteinscript™ II kit, Ambion, Austin, TX; Quick Coupled System, Promega, Madison, WI; Expressway, Invitrogen, Carlsbad, CA). Such systems may incorporate various elements to optimize the efficiency of transcription and translation, such as polyadenylation sequences, consensus ribosomal binding (Kozak) sequences, Shine-Dalgarno sequences and/or other regulatory sequences known in the art.

[0028] In different embodiments of the invention, labeled protein and/or peptides may be purified from the crude *in vitro* translation mixture prior to analysis by the disclosed methods or alternatively may be analyzed without purification.

Protein Expression in Host Cells

[0029] Proteins and/or peptides encoding target protein and/or peptides of interest may be incorporated into expression vectors for transformation into host cells and production of the encoded protein and/or peptides. Non-limiting examples of host cell lines known in the art include bacteria such as *E. coli*, yeast such as *Pichia pastoris*, and mammalian cell lines such as VERO cells, HeLa cells, Chinese hamster ovary cell lines, human embryonic kidney (HEK) 293 cells, mouse neuroblastoma N2A cells, or the W138, BHK, COS-1, COS-7, 293, HepG2, 3T3, RIN, L-929 and MDCK cell lines. These and other host cell lines may be obtained from standard sources, such as the American Type Culture Collection (Rockville, MD) or commercial vendors.

[0030] A complete gene can be expressed or fragments of a gene encoding portions of a protein can be expressed. The gene or gene fragment encoding protein(s) of interest may be inserted into an expression vector by standard cloning techniques. Expression libraries containing part or all of the messenger RNAs expressed in a given cell or tissue type may be prepared by known techniques. Such libraries may be screened for clones encoding particular protein and/or peptides of interest, for example using antibody or oligonucleotide probes and known screening techniques.

[0031] The engineering of DNA segment(s) for expression in a prokaryotic or eukaryotic system may be performed by techniques generally known in the art. Any known expression system may be employed for protein expression. Expression vectors may comprise various known regulatory elements for protein expression, such as promoters, enhancers, ribosome binding sites, termination sequences, polyadenylation sites, *etc.*

[0032] Promoters commonly used in bacterial expression vectors include the β-lactamase, lactose and tryptophan promoter systems. Suitable promoter sequences in yeast expression vectors include the promoters for 3-phosphoglycerate kinase or other glycolytic enzymes. Promoters of use for mammalian cell expression may be derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late

promoter or the early and late promoters of SV40). Many other promoters are known and may be used in the practice of the disclosed methods.

[0033] Eukaryotic expression systems of use include, but are not limited to, insect cell systems infected with, for example, recombinant baculovirus, or plant cell systems infected with recombinant cauliflower mosaic virus or tobacco mosaic virus. In an exemplary insect cell system, *Autographa californica* nuclear polyhidrosis virus is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or the Hi5 cell line (Invitrogen, Carlsbad, CA). Protein and/or peptide coding sequences are cloned into, for example, the polyhedrin gene of the virus under control of the polyhedrin promoter. Recombinant viruses containing the cloned gene are then used to infect *Spodoptera frugiperda* cells and the inserted gene is expressed (e.g., U.S. Patent No. 4,215,051; Kitts *et al.*, Biotechniques 14:810-817, 1993; Lucklow *et al.*, J. Virol., 67:4566-79, 1993). Other exemplary insect cell expression vectors are based on baculovirus vectors, for example, pBlueBac (Invitrogen, Sorrento, CA).

[0034] An exemplary expression system in mammalian cell lines may utilize adenovirus as an expression vector. Coding sequences may be ligated to, *e.g.*, the adenovirus late promoter. The cloned gene may be inserted into the adenovirus genome by *in vitro* or *in vivo* recombination. Insertion in a non-essential region of the viral genome (*e.g.*, region E1 or E3) results in a recombinant virus that is capable of infecting and expressing cloned protein and/or peptides in mammalian host cells. The disclosed examples are not limiting and any known expression vector may be used.

[0035] Expressed protein and/or peptides may be partially or completely purified before analysis. In some embodiments of the invention, protein purification may be facilitated by expressing cloned sequences as fusion protein and/or peptides containing short leader sequences that allow rapid affinity purification. Examples of such fusion protein expression systems are the glutathione S-transferase system (Pharmacia, Piscataway, NJ), the maltose binding protein system (NEB, Beverley, MA), the FLAG system (IBI, New Haven, CT), and the 6xHis system (Qiagen, Chatsworth, CA). In one embodiment of the invention, the leader sequence is linked to a protein by a specific recognition site for a protease. Examples of suitable sequences include those recognized by the Tobacco Etch Virus protease (Life Technologies, Gaithersburg, MD) or

Factor Xa (New England Biolabs, Beverley, MA). Alternatively, expressed protein and/or peptides may be purified by standard techniques discussed above.

Cross-linking agents

[0036] In various embodiments of the invention, cross-linking reagents known in the art, such as homo-bifunctional, hetero-bifunctional and/or photoactivatable cross-linking reagents may be used to link inhibitory peptides to one or more of the BRC enzymes. Non-limiting examples of such reagents include bisimidates; 1,5-difluoro-2,4-(dinitrobenzene); N-hydroxysuccinimide ester of suberic acid; disuccinimidyl tartarate; dimethyl-3,3'-dithio-bispropionimidate; N-succinimidyl-3-(2-pyridyldithio)propionate; 4-(bromoaminoethyl)-2-nitrophenylazide; and 4-azidoglyoxal. These are exemplary only and any cross-linking reagent known in the art may be used within the scope of the invention.

Synthetic Peptides

[0037] The inhibitory peptides of the invention may be synthesized, in whole or in part, in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. See, for example, Stewart and Young, (*Solid Phase Peptide Synthesis*, 2d. ed., Pierce Chemical Co., 1984); Tam *et al.*, (*J. Am. Chem. Soc.*, 105:6442, 1983); Merrifield (*Science*, 232: 341-347, 1986); and Barany and Merrifield (*The Peptides*, Gross and Meienhofer, eds., Academic Press, New York, pp. 1-284, 1979), each incorporated herein by reference. Short peptide sequences, usually from about 6 up to about 35 to 50 amino acids, can be readily synthesized by such methods. Alternatively, recombinant DNA technology may be employed wherein a nucleotide sequence which encodes a peptide of the invention is inserted into an expression vector, transformed or transfected into an appropriate host cell, and cultivated under conditions suitable for expression. Synthetic peptides of any desired sequence that are specific for a selected target protease or other activating protein or peptide may be produced by any such known method.

Detectors

[0038] In various embodiments of the invention, photons generated by BRC may be quantified using a detector, such as a charge coupled device (CCD). Other exemplary detectors

include photodiodes, avalanche photodiodes, photomultiplier tubes, multianode photomultiplier tubes, phototransistors, vacuum photodiodes, silicon photodiodes, and CCD cameras.

[0039] In certain embodiments of the invention, a highly sensitive cooled CCD detector may be used. The cooled CCD detector has a probability of single-photon detection of up to 80%, a high spatial resolution pixel size (5 microns), and sensitivity in the visible through near infrared spectra. (Sheppard, Confocal Microscopy: Basic Principles and System Performance in: Multidimensional Microscopy, P.C. Cheng *et al.* eds., Springer-Verlag, New York, NY pp. 1-51, 1994.) In another embodiment of the invention, a coiled image-intensified coupling device (ICCD) may be used as a photodetector that approaches single-photon counting levels (U.S. Pat. No. 6,147,198). A small number of photons triggers an avalanche of electrons that impinge on a phosphor screen, producing an illuminated image. This phosphor image is sensed by a CCD chip region attached to an amplifier through a fiber optic coupler.

[0040] In some embodiments of the invention, an avalanche photodiode (APD) may be made to detect low light levels. The APD process uses photodiode arrays for electron multiplication effects (U.S. Pat. No. 6,197,503). The invention is not limited to the disclosed embodiments and it is contemplated that any light detector known in the art that is capable of accumulating photons over a time interval may be used in the disclosed methods and apparatus.

[0041] In all of the above embodiments the generated photons from the sample can either reach the detector directly or be guided and/or focused onto the detector by a secondary system such as a number of lenses, reflecting mirror systems, optical waveguides and optical fibers or a combination of those.

EXAMPLES

Example 1: BRC Assay

BRC Protocol

[0042] Methods applicable to the BRC protocol are known in the art. (See, *e.g.*, U.S. Patent Application Serial No. 10/186,455, filed June 28, 2002, the entire text of which is incorporated herein by reference.) In an exemplary embodiment, cDNA synthesis may employ an RNA/primer mixture containing 5 μl total RNA and 1 μl of 0.5 μg/μl oligo(dT) random primer or gene specific primer, incubated at 70°C for 10 min and then placed on ice for at least 1 min. A

reaction mixture containing 2 μl 10X buffer (0.1 M Tris-Acetate pH 7.75, 5 mM EDTA, 50 mM Mg-acetate, 2 mM kinase free dNTP and 0.1 M dithiothreitol) in which dATP is replaced with α-thio dATP may be added to the RNA/primer mixture, mixed gently, collected by brief centrifugation and then incubated at 42°C for 5 min. After addition of 200 U of SuperScript II reverse transcriptase, the tube may be incubated at 40°C for 15 min. The reaction may be terminated by heating at 70°C for 15 min and then chilling on ice. The dNTP used in cDNA synthesis should be kinase free. In preferred embodiments dATP is replaced with alpha-thio dATP or other analogs known in the art that are not good substrates for luciferase (see, e.g. Porter *et al., Nucl. Acids Res.* 25:1611-1617, 1997).

[0043] An aliquot of synthesized cDNA may be added to 50 μl of reaction mixture (see Ronaghi *et al., Anal. Biochem.* 242:84-89, 1996 with modifications) containing 250 ng luciferase (Promega, Madison, WI), 50 mU ATP sulfurylase (Sigma Chemical Co., St. Louis, MO), 2 mM dithiothreitol, 100 mM Tris-Acetate pH 7.75, 0.5 mM EDTA, 0.5 mg BSA, 0.2 mg polyvinylpyrrolidone (M_r 360.000), 10 μg D-luciferin (Biothema, Dalaro, Sweden), 5 mM magnesium acetate and 10 attomole to 0.01 attomole purified pyrophosphate or ATP. The addition of very low amounts of pyrophosphate or ATP (or analogs) is important to decrease background light emission from the reaction mixture. Although the precise mechanism is unknown, BRC performed without adding small amounts of ATP or PPi exhibits background luminescence that precludes accurate measurement of target molecules present in amounts of about a femtomole or lower. Inorganic pyrophosphate present in the cDNA sample as a result of polymerase mediated dNTP incorporation may be converted to ATP by sulfurylase. The ATP may be used to generate light in a luciferin/luciferase reaction.

Modulation of BRC Enzyme Activity with a Peptide Linker

[0044] An exemplary protocol for a protease based BRC assay is illustrated in FIG. 1. A multi-chain amino acid peptide is linked to the luciferase enzyme, thereby inactivating the enzyme activity. A sample containing a target protease is prepared. When the BRC assay (with the inactivated luciferase) is added to the sample. The protease, which is specific to the peptide sequence attached to luciferase, will cleave the peptide thereby activating the luciferase within the solution to generate light. If the protease of interest is not in solution, then there is no cleavage of peptide. The luciferase enzyme remains inactive and no light generated (FIG. 1)

[0045] This method is not limited to assay of proteases. Any molecule or compound that can be linked to luciferase and inactivate the enzyme can also work as long as something in the test sample can restore that activity. Nor is the method limited to modification of luciferase activity. Other BRC enzymes, such as ATP sulfurylase, could also be inhibited by peptide linkage and activated by a target protease or other protein or peptide.

Detection Devices

[0046] The number of the photons generated by BRC may be measured using any known type of photodetector. Common devices that may be used include photodiodes, photomultiplier tubes (PMTs), charge coupled devices (CCDs), and photo-resistive materials. Luciferase-catalyzed photon generation has a quantum yield (Q.E.) of approximately 0.88, with the wavelength maximum depending on the type of luciferase used. For various types of luciferase, that can be anyplace within the visible range of the spectrum. Exemplary embodiments use firefly luciferase, which has a maximum intensity at 562 nm.

[0047] The photosensitive device is typically either in direct proximity of the BRC reaction to directly receive incident photons, or relatively far from the buffer with a light coupling device (*e.g.* optical fiber or mirror system) capable of directing light from the sample to the detector. In an exemplary embodiment, a UDT-PIN-UV-50-9850-1 photodiode (Hamamatsu Corp., Hamamatsu, Japan) may be used with a transimpedance amplifier with a gain of 10^8 volts/amp.

* * *

[0048] All of the COMPOSITIONS, METHODS and APPARATUS disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the COMPOSITIONS, METHODS and APPARATUS and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications

apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

WHAT IS CLAIMED IS:

1. A method comprising:

a) obtaining at least one sample suspected of containing one or more target proteins and/or peptides;

b) obtaining a BRC assay mixture with at least one BRC enzyme inactivated by peptide linkage;

c) exposing the sample to the BRC assay mixture;

d) activating the inactivated BRC enzyme;

e) generating pyrophosphate (PPi); and

f) producing light by a bioluminescence regenerative cycle (BRC).

2. The method of claim 1, further comprising detecting the target protein and/or peptide.

3. The method of claim 1, further comprising identifying the the target protein and/or peptide.

4. The method of claim 1, further comprising quantifying the target protein and/or peptide.

5. The method of claim 1, further comprising accumulating the number of photons produced over a time interval.

6. The method of claim 5, further comprising determining the number of target proteins and/or peptides in the sample.

7. The method of claim 1, wherein the target protein is a protease and said protease removes the linked peptide from the BRC enzyme.

8. The method of claim 1, wherein the inactivated BRC enzyme is luciferase or ATP sulfurylase.

9. The method of claim 8, further comprising measuring protein-protein binding.

10. The method of claim 1, further comprising measuring gene expression levels in a sample from a cell line, tissue, organ or subject.

11. The method of claim 10, wherein the subject is a plant, animal, human, microorganism and/or bacterium.

12. The method of claim 11, further comprising measuring the expression of a single gene.

13. The method of claim 11, further comprising measuring the expression of two or more genes.

14. The method of claim 1, further comprising detecting a pathogen in the sample.

15. The method of claim 14, wherein the pathogen is selected from the group consisting of a virus, bacterium, *Salmonella*, *Streptococcus*, *Legionella*, *E. coli*, *Giardia*, *Cryptosporidium*, *Rickettsia*, spore, mold, yeast, algae, amoebae, dinoflagellate, unicellular organism, parasite, nematode, trypanosome, or a diseased cell.

16. The method of claim 15, further comprising detecting a family of pathogens in the sample.

17. The method of claim 1, wherein the bioluminescence regenerative cycle utilizes adenosine 5'-phosphosulphate (APS), ATP sulfurylase, luciferin and luciferase.

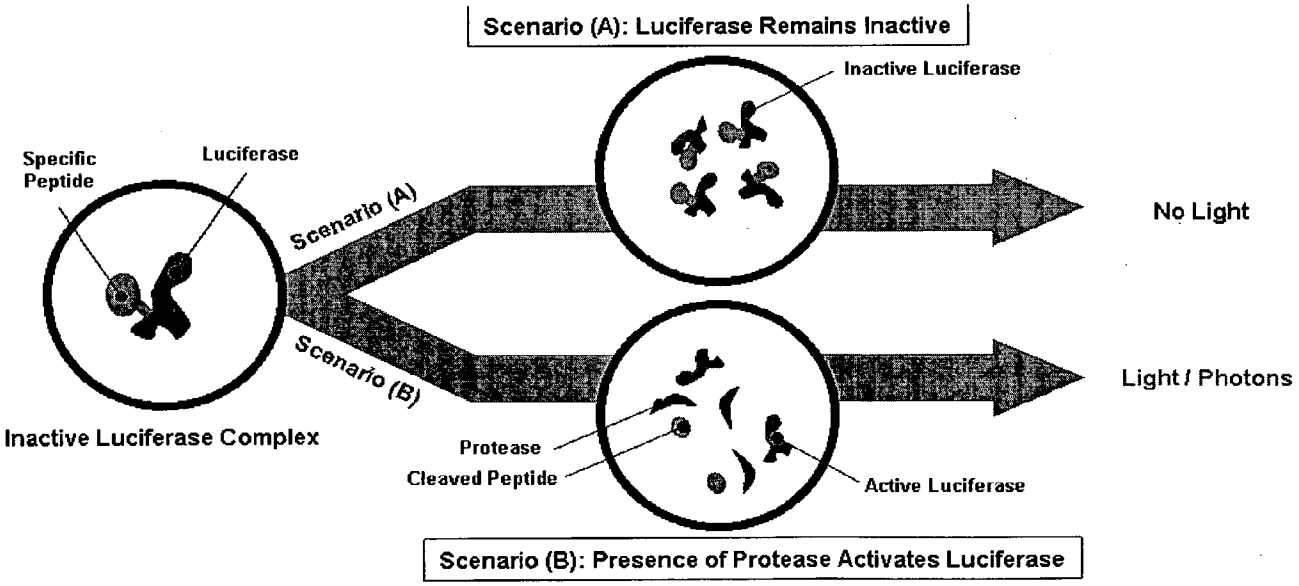
18. The method of claim 17, further comprising adding ATP or PPi to the sample before light is produced.

19. The method of claim 18, wherein the amount of ATP or PPi is between 0.01 and 10 attomoles.

ABSTRACT

[0049] The present invention concerns methods of quantifying proteins and/or peptides using a bioluminescence regenerative cycle (BRC). In BRC, steady state levels of bioluminescence result from processes that produce pyrophosphate. Pyrophosphate reacts with APS in the presence of ATP sulfurylase to produce ATP. The ATP reacts with luciferin in a luciferase-catalyzed reaction, producing light and regenerating pyrophosphate. The pyrophosphate is recycled to produce ATP and the regenerative cycle continues. Because the kinetic properties of ATP sulfurylase are much faster than luciferase, a steady state results wherein concentrations of ATP and pyrophosphate and the rate of light production remain relatively constant. Photons are counted over a time interval to determine the number of target molecules present in the initial sample. According to the present invention, one or more of the BRC enzymes is inhibited by linkage to a peptide. The presence of a target protein and/or peptide in a sample activates the BRC enzyme and allows light production. In some cases, activation involves proteolytic removal of the inhibitory peptide.

FIG. 1


Scenario (A): Luciferase Remains Inactive
Inactive Luciferase
Specific Peptide
Luciferase
Scenario (A)
No Light
Scenario (B)
Light / Photons
Inactive Luciferase Complex
Protease
Cleaved Peptide
Active Luciferase
Scenario (B): Presence of Protease Activates Luciferase

PATENT

Attorney Docket: 005852.P007Z

APPLICATION FOR UNITED STATES LETTERS PATENT

for

CLONED SEQUENCE INSERT DETECTION USING BRC (BIOLUMINESCENCE REGENERATIVE CYCLE)

by

Nader Pourmand

Arjang Hassibi

and

Siavash Ghazvini

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP
12400 Wilshire Boulevard, Seventh Floor
Los Angeles, California 90025-1030

Telephone: (303) 740-1980
Facsimile: (303) 740-6962

EXPRESS MAIL CERTIFICATE OF MAILING
"Express Mail" Mailing No.: EL 880593806 US

I hereby certify that I am causing the above-referenced correspondence to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated below and that this paper or fee has been addressed to Box Provisional Application, Assistant Commissioner for Patents, Washington, D.C. 20231.

Date of Deposit: December 20, 2002

Name of Person Mailing Correspondence: Lori E. True

Lori E. True
Signature

December 20, 2002
Date

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates to the field of nucleic acid detection and/or quantification. More particularly, the present invention concerns novel approaches to detection and/or quantification of cloned sequences that have been inserted into a plasmid, cosmid, phage, YAC, BAC or any other known type of vector.

Description of Related Art

[0002] Methods of precise and highly sensitive detection and/or quantification of cloned nucleic acids are of use for a variety of medical, forensic, epidemiological, public health, biological warfare and other applications. A variety of molecular biology and genomic techniques would benefit from the availability of precise and sensitive methods for nucleic acid detection and/or quantification.

[0003] DNA microarrays provide a platform for detecting and identifying nucleic acids by hybridization with sequence specific oligonucleotide probes attached to chips in precise arrays. (*E.g.*, Schena *et al.*, *Science* 270:467-470, 1995; Shalon *et al.*, *Genome Res.* 6:639-645, 1996; Pease *et al.*, *Proc. Natl. Acad. Sci. USA* 91:5022-26, 1994). Microarray technology is an extension of previous hybridization-based methods, such as Southern and Northern blotting, that have been used to identify and quantify nucleic acids in biological samples (Southern, *J. Mol. Biol.* 98:503-17, 1975; Pease *et al.*, *Proc. Natl. Acad. Sci. USA* 93:10614-19, 1996). Identification of a target nucleic acid in a sample typically involves fluorescent detection of the nucleic acid hybridized to an oligonucleotide at a particular location on the array. Fluorescent detection is too insensitive to detect very low levels of a target nucleic acid in a sample. It is also more qualitative than quantitative. More accurate and sensitive methods for nucleic acid quantification are needed.

[0004] Real time PCR™ (polymerase chain reaction) is another technique for which accurate and sensitive detection and/or quantification are needed (*e.g.*, Model 770 TaqMan® system, Applied Biosystems, Foster City, CA). Typically, if the target of interest is present, it will be amplified by replication using flanking primers and a nucleic acid polymerase. A probe, which may consist of a complementary oligonucleotide with attached reporter and quencher dyes, is

designed to bind to the amplified target nucleic acid between the two primer-binding sites. The nuclease activity of the polymerase cleaves the probe, resulting in an increase in fluorescence of the reporter dye after it is separated from the quencher. PCR based fluorescence detection of target nucleic acids is more sensitive, due to the amplification effect of the technique. However, detection and/or quantification of the target may be complicated by a variety of factors, such as contaminating nuclease activity or variability in the efficiency of amplification.

[0005] Single nucleotide polymorphisms (SNPs) are of increasing interest in molecular biology, genomics and disease diagnostics. SNP detection may be used for haplotype construction in genetic studies to identify and/or detect genes associated with various disease states, as well as drug sensitivity or resistance. SNPs may be detected by a variety of techniques, such as DNA sequencing, fluorescence detection, mass spectrometry or DNA microarray hybridization (*e.g.*, U.S. Patent Nos. 5,885,775; 6,368,799). Existing methods of SNP detection may suffer from insufficient sensitivity or an unacceptably high level of false positive and/or false negative results. A need exists for more sensitive and accurate methods of detecting SNPs.

[0006] Pyrophosphate based detection systems have been used for DNA sequencing (*e.g.*, Nyren and Lundin, *Anal. Biochem.* 151:504-509, 1985; U.S. Patent Nos. 4,971,903; 6,210,891; 6,258,568; 6,274,320, each incorporated herein by reference). The method uses a coupled reaction wherein pyrophosphate is generated by an enzyme-catalyzed process, such as nucleic acid polymerization. The pyrophosphate is used to produce ATP, in an ATP sulfurylase catalyzed reaction with adenosine 5'-phosphosulphate (APS). The ATP in turn is used for the production of light in a luciferin-luciferase coupled reaction. The present invention provides a novel method of pyrophosphate-based detection for use in detecting, identifying and/or quantifying cloned nucleic acid sequences.

SUMMARY OF THE INVENTION

[0007] The present invention fulfills an unresolved need in the art by providing methods for accurately detecting, identifying and/or quantifying nucleic acids sequences that have been inserted into a plasmid or other known vector. The disclosed methods provide increased sensitivity and accuracy of target molecule detection, identification and/or quantification

compared to prior art methods. The disclosed methods are generally referred to as "bioluminescence regenerative cycle" or BRC.

[0008] In certain embodiments of the invention, the methods may comprise obtaining at least one sample suspected of containing one or more target nucleic acids, generating pyrophosphate by replication of the target nucleic acid, producing light by a bioluminescence regenerative cycle, accumulating the total number of photons produced over different time intervals, comparing the photon accumulation with the background photon emission and determining the number of target nucleic acids in the sample. In particular embodiments, the target nucleic acid is replicated by polymerase chain reaction amplification, although in alternative embodiments any process or reaction that results in the production of pyrophosphate may be coupled to BRC analysis.

[0009] In various embodiments of the invention, the disclosed methods are of use for a wide variety of applications for which nucleic acid detection, identification and/or quantification is desired. Such applications include, but are not limited to, measuring gene expression levels, detecting and/or quantifying pathogens in a sample, performing real-time PCR™ analysis and detecting single nucleotide polymorphisms (SNPs). In particular embodiments, the BRC method may comprise a rolling circle method of nucleic acid replication.

[0010] In particular embodiments of the invention, pyrophosphate is generated by a reaction that generates pyrophosphate (PPi), such as PCR, transcription and/or DNA replication. In preferred embodiments, sequence specific primers are used to limit replication to a particular target nucleic acid in the sample. The sequence specific primers are designed to not bind to other nucleic acids that may be present in the sample. In more preferred embodiments, the pyrophosphate producing reaction is allowed to proceed to completion before BRC analysis. Once the reaction is complete, the pyrophosphate is reacted with APS in the presence of ATP sulfurylase to produce ATP and sulphate. The ATP is reacted with oxygen and luciferin in the presence of luciferase to yield oxyluciferin, AMP and pyrophosphate. For each molecule of pyrophosphate that is cycled through BRC, a photon of light is emitted and one molecule of pyrophosphate is regenerated. Because of the relative kinetic rates of luciferase and ATP sulfurylase, a steady state is reached in which the concentrations of ATP and pyrophosphate and the level of photon output remain relatively constant over an extended period of time. The number of photons may be counted

over a time interval to determine the number of target nucleic acids in the sample. The very high sensitivity of BRC is related in part to the integration of light output over time, in contrast to other methods that measure light output at a single time point or at a small number of fixed time points. The ability to vary the length of time over which photon integration occurs also contributes to the very high dynamic range for nucleic acid molecule quantification. The detection noise is also significantly reduced by increasing the length of integration.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[0012] FIG. 1 illustrates the use of BRC to detect complex genomic DNA inserted into a plasmid vector, with and without amplification of the target nucleic acid sequence. Detection and quantification of a target RO 52 insert sequence was demonstrated.

DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

Definitions

[0013] Terms that are not otherwise defined herein are used in accordance with their plain and ordinary meaning.

[0014] As used herein, "a" or "an" may mean one or more than one of an item.

[0015] As used herein, "luminescence" refers to the emission of light that does not derive energy from the temperature of the emitting body (i.e., emission of light other than incandescent light). "Luminescence" includes, but is not limited to, fluorescence, phosphorescence, thermoluminescence, chemiluminescence, electroluminescence and bioluminescence. "Luminescent" refers to an object that exhibits luminescence. In preferred embodiments, the light is in the visible spectrum. However, the present invention is not limited to visible light, but includes electromagnetic radiation of any frequency.

[0016] "Nucleic acid" means either DNA, RNA, single-stranded, double-stranded or triple stranded and any chemical modifications thereof. Virtually any modification of the nucleic acid

is contemplated by this invention. "Nucleic acid" encompasses, but is not limited to, oligonucleotides and polynucleotides. Within the practice of the present invention, a "nucleic acid" may be of any length.

BRC Detection Method

[0017] Various embodiments of the invention concern novel methods for detecting, identifying and/or quantifying nucleic acid molecules without labeling of any target, capture or probe molecules. Such label free methods are advantageous with respect to sensitivity, expense and ease of use. The BRC methods involve the luminescent detection of pyrophosphate (PPi) molecules released from an enzyme-catalyzed reaction, such as RNA and/or DNA polymerization. As part of the technique, a bioluminescence regenerative cycle (BRC) is triggered by the release of inorganic pyrophosphate (PPi) from polymerization of a nucleic acid.

[0018] The BRC regenerative cycle can be utilized with any reaction that generates pyrophosphate (PPi), such as nucleic acid polymerization. The PPi generated reacts with APS, catalyzed by ATP-sulfurylase enzyme, which results in the production of ATP and inorganic sulphate. In another reaction, luciferin and luciferase consume ATP as an energy source to generate light, AMP and oxyluciferin and to regenerate PPi. Thus, after each BRC cycle, a quantum of light is generated for each molecule of PPi in solution, while the net concentration of ATP in solution remains relatively stable and is proportional to the initial concentration of PPi. In the course of the reactions, APS and luciferin are consumed and AMP and oxyluciferin are generated, while ATP sulfurylase and luciferase remain constant. The invention is not limited as to the type of luciferase used. Although certain disclosed embodiments utilized firefly luciferase, any luciferase known in the art may be used in the disclosed methods.

[0019] As a result of the BRC process, the photon emission rate remains steady and is a monotonic function of the amount of PPi in the initial mixture. For very low concentrations of PPi (10^{-8} M or less), the total number of photons generated in a fixed time interval is proportional to the number of PPi molecules. Where PPi is generated by the polymerase catalyzed replication of a target nucleic acid, the number of photons generated in a fixed time interval is proportional to the quantity of the target nucleic acid present in the sample.

[0020] The basic concept of enzymatic light generation from PPi molecules was introduced almost two decades ago (Nyren and Lundin, 1985; Nyren, *Anal. Biochem.* 167:235-238, 1987).

Pyrophosphate based luminescence has been used for DNA sequencing (Ronaghi *et al.*, *Anal. Biochem.* 242:84-89, 1996) and SNP detection (Nyren *et al.*, *Anal. Biochem.* 244:367-373, 1997). The present methods provide additional procedures for accurately detecting, identifying and/or quantifying specific target nucleic acids inserted into one or more vectors, in the presence of contaminants and detector noise. The novel system and methods have an intrinsic controllable dynamic range up to seven orders of magnitude and are sensitive enough to detect target nucleic acids at attomole (10^{-18}) or lower levels.

Enzymatic Bioluminescence Cycle

[0021] To generate photons from pyrophosphate, ATP-sulfurylase (Ronesto *et al.*, *Arch. Biochem. Biophys.* 290:66-78, 1994; Beynon *et al. Biochemistry*, 40, 14509-14517, 2001) is used to catalyze the transfer of the adenylyl group from APS to PPi, producing ATP and inorganic sulfate.

[0022] Next, luciferase catalyzes the slow consumption of ATP, luciferin and oxygen to generate a single photon (λ_{max}=562nm, Q.E. $\approx$ 0.88) per ATP molecule, regenerating a molecule of PPi and producing AMP, CO_2 and oxyluciferin (Eq. 5). (Brovko *et al.*, *Biochem.* (Moscow) 59:195-201, 1994) Because the luciferase reaction is significantly slower than the ATP-sulfurylase reaction, in the presence of sufficient amounts of the substrates APS and luciferin a steady state cycle should be maintained, in which the concentration of ATP and the resulting levels of light emission remain relatively constant for a considerable time.

[0023] Because the steady-state photon emission is proportional to the initial concentration of PPi, the presence of minute amounts of PPi produced by a polymerase or other reaction should result in a detectable shift in baseline luminescence, even in the presence of considerable amounts of noise. The number of photons generated over time by the BRC cycle can potentially be orders of magnitude higher than the initial number of PPi molecules, which makes the system extremely sensitive compared to prior art methods of nucleic acid quantification. The increased sensitivity is provided by having a time-dependent amplification of light emission for each molecule of PPi present at the start of the BRC cycle.

Nucleic Acids

[0024] Samples comprising target nucleic acids inserted into vectors may be prepared by any technique known in the art. In certain embodiments, the analysis may be performed on crude sample extracts, containing complex mixtures of nucleic acids, proteins, lipids, polysaccharides and other compounds. Such samples are likely to contain contaminants that could potentially interfere with the BRC process. In preferred embodiments, the vector nucleic acids and/or inserts may be partially or fully separated from other sample constituents before initiating the BRC analysis.

[0025] Methods for partially or fully purifying nucleic acids from complex mixtures, such as cell homogenates or extracts, are well known in the art. (See, *e.g.*, Guide to Molecular Cloning Techniques, eds. Berger and Kimmel, Academic Press, New York, NY, 1987; Molecular Cloning: A Laboratory Manual, 2nd Ed., eds. Sambrook, Fritsch and Maniatis, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1989). Generally, cells, tissues or other source material containing nucleic acids are first homogenized, for example by freezing in liquid nitrogen followed by grinding in a mortar and pestle. Certain tissues may be homogenized using a Waring blender, Virtis homogenizer, Dounce homogenizer or other homogenizer. Crude homogenates may be extracted with detergents, such as sodium dodecyl sulphate (SDS), Triton X-100, CHAPS (3-[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate), octylglucoside or other detergents known in the art. As is well known, nuclease inhibitors such as RNase or DNase inhibitors may be added to prevent degradation of target nucleic acids.

[0026] Extraction may also be performed with chaotrophic agents such as guanidinium isothiocyanate, or organic solvents such as phenol. In some embodiments, protease treatment, for example with proteinase K, may be used to degrade cell proteins. Particulate contaminants may be removed by centrifugation or ultracentrifugation. Dialysis against aqueous buffer of low ionic strength may be of use to remove salts or other soluble contaminants. Nucleic acids may be precipitated by addition of ethanol at -20°C, or by addition of sodium acetate (pH 6.5, about 0.3 M) and 0.8 volumes of 2-propanol. Nucleic acids may be collected by centrifugation or other known methods, such as preparative agarose gel electrophoresis or use of a commercially available column for preparation of plasmids or other vectors. The skilled artisan will realize

that the procedures listed above are exemplary only and that many variations may be used, depending on the particular type of nucleic acid to be analyzed.

[0027] In certain embodiments, the target nucleic acids of interest may be prepared as part of a library of insert-containing vectors, such as BAC, YAC, cosmid, plasmid or phage libraries containing nucleic acid inserts. (See, *e.g.*, Berger and Kimmel, 1987; Sambrook *et al.*, 1989.)

Methods of Immobilization

[0028] In various embodiments, the nucleic acids to be analyzed may be attached to a solid surface (or immobilized). Immobilization of nucleic acids may be achieved by a variety of methods involving either non-covalent or covalent attachment between the nucleic acid and the surface. In an exemplary embodiment, immobilization may be achieved by coating a surface with streptavidin or avidin and the subsequent attachment of a biotinylated polynucleotide (Holmstrom *et al.*, *Anal. Biochem.* 209:278-283, 1993). Immobilization may also occur by coating a silicon, glass or other surface with poly-L-Lys (lysine), followed by covalent attachment of either amino- or sulfhydryl-modified nucleic acids using bifunctional crosslinking reagents (Running *et al.*, *BioTechniques* 8:276-277, 1990; Newton *et al.*, *Nucleic Acids Res.* 21:1155-62, 1993). Amine residues may be introduced onto a surface through the use of aminosilane for cross-linking.

[0029] Immobilization may take place by direct covalent attachment of 5'-phosphorylated nucleic acids to chemically modified surfaces (Rasmussen *et al.*, *Anal. Biochem.* 198:138-142, 1991). The covalent bond between the nucleic acid and the surface is formed by condensation with a water-soluble carbodiimide. This method facilitates a predominantly 5'-attachment of the nucleic acids via their 5'-phosphates.

[0030] DNA is commonly bound to glass by first silanizing the glass surface, then activating with carbodiimide or glutaraldehyde. Alternative procedures may use reagents such as 3-glycidoxypropyltrimethoxysilane (GOP) or aminopropyltrimethoxysilane (APTS) with DNA linked *via* amino linkers incorporated either at the 3′ or 5′ end of the molecule. DNA may be bound directly to membrane surfaces using ultraviolet radiation. Other non-limiting examples of immobilization techniques for nucleic acids are disclosed in U.S. Patent Nos. 5,610,287, 5,776,674 and 6,225,068.

[0031] The type of surface to be used for immobilization of the nucleic acid is not limiting. In various embodiments, the immobilization surface may be magnetic beads, non-magnetic beads, a planar surface, or any other conformation of solid surface comprising almost any material, so long as the material is sufficiently durable and inert to allow the BRC process to occur. Non-limiting examples of surfaces that may be used include glass, silica, silicate, PDMS, silver or other metal coated surfaces, nitrocellulose, nylon, activated quartz, activated glass, polyvinylidene difluoride (PVDF), polystyrene, polyacrylamide, other polymers such as poly(vinyl chloride), poly(methyl methacrylate) or poly(dimethyl siloxane), and photopolymers which contain photoreactive species such as nitrenes, carbenes and ketyl radicals capable of forming covalent links with nucleic acids (See U.S. Pat. Nos. 5,405,766 and 5,986,076).

[0032] Bifunctional cross-linking reagents may be of use in various embodiments, such as attaching a nucleic acid to a surface. The bifunctional cross-linking reagents can be divided according to the specificity of their functional groups, *e.g.*, amino, guanidino, indole, or carboxyl specific groups. Of these, reagents directed to free amino groups are popular because of their commercial availability, ease of synthesis and the mild reaction conditions under which they can be applied. Exemplary methods for cross-linking molecules are disclosed in U.S. Patent Nos. 5,603,872 and 5,401,511. Cross-linking reagents include glutaraldehyde (GAD), bifunctional oxirane (OXR), ethylene glycol diglycidyl ether (EGDE), and carbodiimides, such as 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC).

Polymerases

[0033] In certain embodiments, the disclosed methods may involve binding of a DNA polymerase to a primer molecule and the catalyzed addition of nucleotide precursors to the 3' end of a primer. Non-limiting examples of polymerases of potential use include DNA polymerases, RNA polymerases, reverse transcriptases, and RNA-dependent RNA polymerases. The differences between these polymerases in terms of their requirement or lack of requirement for primers and promoter sequences are known in the art.

[0034] Non-limiting examples of polymerases that may be of use include *Thermatoga maritima* DNA polymerase, AmplitaqFS™ DNA polymerase, Taquenase™ DNA polymerase, ThermoSequenase™, Taq DNA polymerase, Qbeta™ replicase, T4 DNA polymerase, *Thermus thermophilus* DNA polymerase, RNA-dependent RNA polymerase and SP6 RNA polymerase.

Commercially available polymerases including Pwo DNA Polymerase from Boehringer Mannheim Biochemicals (Indianapolis, IN); Bst Polymerase from Bio-Rad Laboratories (Hercules, CA); IsoTherm™ DNA Polymerase from Epicentre Technologies (Madison, WI); Moloney Murine Leukemia Virus Reverse Transcriptase, *Pfu* DNA Polymerase, Avian Myeloblastosis Virus Reverse Transcriptase, *Thermus flavus (Tfl)* DNA Polymerase and *Thermococcus litoralis (Tli)* DNA Polymerase from Promega (Madison, WI); RAV2 Reverse Transcriptase, HIV-1 Reverse Transcriptase, T7 RNA Polymerase, T3 RNA Polymerase, SP6 RNA Polymerase, RNA Polymerase E. coli, *Thermus aquaticus* DNA Polymerase, T7 DNA Polymerase +/- 3'→5' exonuclease, Klenow Fragment of DNA Polymerase I, Thermus 'ubiquitous' DNA Polymerase, and DNA polymerase I from Amersham Pharmacia Biotech (Piscataway, NJ).

[0035] Methods of using polymerases and compositions suitable for use in such methods are well known in the art (*e.g.,* Berger and Kimmel, 1987; Sambrook *et al.*, 1989).

Primers

[0036] Where primers are required to initiate polymerase activity, they may be obtained by any method known in the art. Generally, primers are between ten and twenty bases in length, although longer primers may be employed. In certain embodiments, primers are designed to be exactly complementary in sequence to a known portion of a target nucleic acid. Methods for synthesis of primers of any sequence are known, for example using an automated nucleic acid synthesizer employing phosphoramidite chemistry. Such instruments may be obtained from commercial sources, such as Applied Biosystems (Foster City, CA) or Millipore Corp. (Bedford ,MA).

Detectors

[0037] In various embodiments of the invention, photons generated by BRC may be quantified using a detector, such as a charge coupled device (CCD). Other exemplary detectors include photodiodes, avalanche photodiodes, photomultiplier tubes, multianode photomultiplier tubes, phototransistors, vacuum photodiodes, silicon photodiodes, and CCD cameras.

[0038] In certain embodiments of the invention, a highly sensitive cooled CCD detector may be used. The cooled CCD detector has a probability of single-photon detection of up to 80%, a high

spatial resolution pixel size (5 microns), and sensitivity in the visible through near infrared spectra. (Sheppard, Confocal Microscopy: Basic Principles and System Performance in: Multidimensional Microscopy, P.C. Cheng *et al.* eds., Springer-Verlag, New York, NY pp. 1-51, 1994.) In another embodiment of the invention, a coiled image-intensified coupling device (ICCD) may be used as a photodetector that approaches single-photon counting levels (U.S. Pat. No. 6,147,198). A small number of photons triggers an avalanche of electrons that impinge on a phosphor screen, producing an illuminated image. This phosphor image is sensed by a CCD chip region attached to an amplifier through a fiber optic coupler.

[0039] In some embodiments of the invention, an avalanche photodiode (APD) may be made to detect low light levels. The APD process uses photodiode arrays for electron multiplication effects (U.S. Pat. No. 6,197,503). The invention is not limited to the disclosed embodiments and it is contemplated that any light detector known in the art that is capable of accumulating photons over a time interval may be used in the disclosed methods and apparatus.

[0040] In all of the above embodiments the generated photons from the sample can either reach the detector directly or be guided and/or focused onto the detector by a secondary system such as a number of lenses, reflecting mirror systems, optical waveguides and optical fibers or a combination of those.

EXAMPLES

Example 1: BRC Assay

Sample Preparation

[0041] Methods applicable to the BRC protocol are disclosed, for example, in U.S. Patent Application Serial No. 10/186,455, filed June 28, 2002, the entire text of which is incorporated herein by reference. In an exemplary embodiment where the nucleic acid insert is transcribed as a messenger RNA, cDNA synthesis may employ an RNA/primer mixture containing 5 μl total RNA and 1 μl of 0.5 μg/μl oligo(dT) random primer or gene specific primer, incubated at 70°C for 10 min and then placed on ice for at least 1 min. A reaction mixture containing 2 μl 10X buffer (0.1 M Tris-Acetate pH 7.75, 5 mM EDTA, 50 mM Mg-acetate, 2 mM kinase free dNTP and 0.1 M dithiothreitol) in which dATP is replaced with α-thio dATP may be added to the RNA/primer mixture, mixed gently, collected by brief centrifugation and then incubated at 42°C for 5 min. After addition of 200 U of SuperScript II reverse transcriptase, the tube may be

incubated at 40°C for 15 min. The reaction may be terminated by heating at 70°C for 15 min and then chilling on ice. The dNTP used in cDNA synthesis should be kinase free. In preferred embodiments dATP is replaced with alpha-thio dATP or other analogs known in the art that are not good substrates for luciferase (see, *e.g.* Porter *et al., Nucl. Acids Res.* 25:1611-1617, 1997).

[0042] An aliquot of synthesized cDNA may be added to 50 μl of reaction mixture (see Ronaghi *et al., Anal. Biochem.* 242:84-89, 1996 with modifications) containing 250 ng luciferase (Promega, Madison, WI), 50 mU ATP sulfurylase (Sigma Chemical Co., St. Louis, MO), 2 mM dithiothreitol, 100 mM Tris-Acetate pH 7.75, 0.5 mM EDTA, 0.5 mg BSA, 0.2 mg polyvinylpyrrolidone (M_r 360,000), 10 μg D-luciferin (Biothema, Dalaro, Sweden), 5 mM magnesium acetate and 10 attomole to 0.01 attomole purified pyrophosphate or ATP. The addition of very low amounts of pyrophosphate or ATP (or analogs) is important to decrease background light emission from the reaction mixture. Although the precise mechanism is unknown, BRC performed without adding small amounts of ATP or PPi exhibits background luminescence that precludes accurate measurement of target molecules present in amounts of about a femtomole or lower. Inorganic pyrophosphate present in the cDNA sample as a result of polymerase mediated dNTP incorporation may be converted to ATP by sulfurylase. The ATP may be used to generate light in a luciferin/luciferase reaction.

Detection Devices

[0043] The number of the photons generated by BRC may be measured using any known type of photodetector. Common devices that may be used include photodiodes, photomultiplier tubes (PMTs), charge coupled devices (CCDs), and photo-resistive materials. Luciferase-catalyzed photon generation has a quantum yield (Q.E.) of approximately 0.88, with the wavelength maximum depending on the type of luciferase used. For various types of luciferase, that can be anyplace within the visible range of the spectrum. Exemplary embodiments use firefly luciferase, which has a maximum intensity at 562 nm.

[0044] The photosensitive device is typically either in direct proximity of the BRC reaction to directly receive incident photons, or relatively far from the buffer with a light coupling device (*e.g.* optical fiber or mirror system) capable of directing light from the sample to the detector. In an exemplary embodiment, a UDT-PIN-UV-50-9850-1 photodiode (Hamamatsu Corp., Hamamatsu, Japan) may be used with a transimpedance amplifier with a gain of 10^8 volts/amp.

Example 2: Cloned Sequence Insert Detection Using BRC

[0045] The BRC procedure may be used to detect a given sequence of nucleic acid that has been inserted into a plasmid or other vector. In exemplary embodiments, a sequence of interest may be amplified from genomic DNA and purified. The amplified product is cloned into a plasmid and transfected into a bacteria. The bacterial is spread on an agar plate and allowed to incubate for a period of time and colonies are formed. A sample of a colony is picked and added to a tube containing buffer. The tube is heated for 5 minutes at 95°C and then cooled.

BRC Analysis With Isothermal Amplification

[0046] BRC assay reagents and isothermal/thermal amplification reagents are added together into the tube with target sequence specific primer(s) and amplified at the appropriate temperature. Light intensity is measured for presence or absence of target sequence.

BRC Analysis With PCR

[0047] PCR reaction mixture was added to the tube contents along with a primer specific to the target sequence. The sample was subjected to one or more cycles of PCR amplification, for example at 95°C (1min), 55°C (1min), and 72°C (1min). In an illustrative embodiment, PCR amplification was performed for 0, 10, 20 or 30 cycles using a RO 52 sequence inserted into a plasmid vector. An aliquot was added to BRC reagents and light intensity was measured for presence or absence of target sequence.

Results

[0048] FIG. 1 shows that BRC can be used with a complex genomic background with and without amplification steps. Bacterial colonies containing a RO 52 sequence inserted into a standard plasmid vector were grown on an agar plate. A colony with less than 100,000 bacteria was isolated and placed into 4 tubes containing buffer. The tubes were heated to 95°C for 5 minutes and then the master mix containing Taq Polymerase, dNTP, primers specific to RO 52 and Mg (2.5 mM $MgCl_2$) was added. Tube 1 was heated to 95°C for 1 min, 55°C for 1 min and 72°C for 1 minute for one cycle. Tubes 2, 3 and 4 were subjected to similar temperature cycles but respectively for 10 cycles, 20 cycles, and 30 cycles. An aliquot of each was added to the BRC assay for PPi measurement. The target RO 52 insert sequence could be detected and quantified in each tube after zero, 10, 20 and 30 cycles of amplification. A reference sample

containing all reagents and biological substances except the RO 52 specific primers showed no detectable signal (FIG. 1).

* * *

[0049] All of the COMPOSITIONS, METHODS and APPARATUS disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the COMPOSITIONS, METHODS and APPARATUS and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

WHAT IS CLAIMED IS:

1. A method comprising:

   a) obtaining at least one sample suspected of containing one or more target nucleic acids inserted into a vector;

   b) generating pyrophosphate (PPi) by replication of the target nucleic acid;

   c) producing light by a bioluminescence regenerative cycle (BRC); and

   e) detecting the target nucleic acid.

2. The method of claim 1, further comprising identifying the target nucleic acid.

3. The method of claim 1, further comprising determining the amount of target nucleic acid in the sample.

4. The method of claim 1, wherein the target nucleic acid comprises genomic DNA.

5. The method of claim 1, wherein the target nucleic acid comprises cDNA.

6. The method of claim 1, wherein the target nucleic acid comprises a single nucleotide polymorphism (SNP) site.

7. The method of claim 6, further comprising amplifying the target DNA using a primer specific for the SNP sequence.

8. The method of claim 1, wherein the target nucleic acid is replicated by polymerase chain reaction (PCR) amplification.

9. The method of claim 1, wherein the target nucleic acid is replicated by a rolling circle technique.

10. The method of claim 1, further comprising measuring gene expression levels in a sample from a cell line, tissue, organ or subject.

11. The method of claim 11, further comprising measuring the expression of two or more genes.

12. The method of claim 1, further comprising isolating total DNA from a sample.

13. The method of claim 12, further comprising restricting the DNA into fragments and inserting the fragments into a vector.

14. Thc method of claim 13, further comprising detecting a pathogen DNA sequence.

15. The method of claim 14, wherein the pathogen is selected from the group consisting of a virus, bacterium, *Salmonella*, *Streptococcus*, *Legionella*, *E. coli*, *Giardia*, *Cryptosporidium*, *Rickettsia*, spore, mold, yeast, algae, amoebae, dinoflagellate, unicellular organism, parasite, nematode, trypanosome, or a diseased cell

16. The method of claim 1, further comprising isolating messenger RNA (mRNA) from a sample.

17. The method of claim 15, further comprising converting the mRNA into double stranded complementary DNA (cDNA).

18. The method of claim 16, further comprising inserting the double stranded cDNA into a vector.

19. The method of claim 1, wherein the vector is a plasmid, cosmid, phage, virus, BAC or YAC vector.

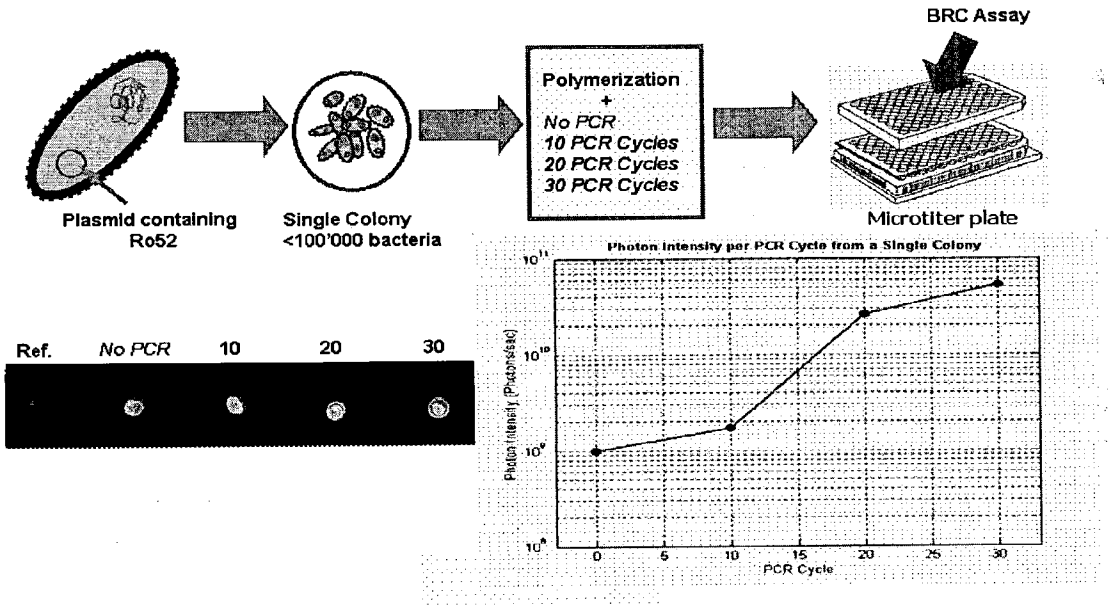
20. The method of claim 1, wherein the bioluminescence regenerative cycle utilizes adenosine 5'-phosphosulphate (APS), ATP sulfurylase, luciferin and luciferase.

21. The method of claim 20, further comprising adding ATP or PPi to the sample before light is produced.

ABSTRACT

[0050] The present invention concerns methods of detecting, identifying and/or quantifying nucleic acids inserted into a vector. In various embodiments, a bioluminescence regenerative cycle (BRC) is used. In BRC, steady state levels of bioluminescence result from processes that produce pyrophosphate. Pyrophosphate reacts with APS in the presence of ATP sulfurylase to produce ATP. The ATP reacts with luciferin in a luciferase-catalyzed reaction, producing light and regenerating pyrophosphate. The pyrophosphate is recycled to produce ATP and the regenerative cycle continues. During the course of the cycle a steady state is achieved wherein concentrations of ATP and pyrophosphate and the rate of light production remain relatively constant. In preferred embodiments, photons are counted over a time interval to determine the number of target molecules present in the initial sample.

FIG. 1


BRC Assay
Polymerization
+
No PCR
10 PCR Cycles
20 PCR Cycles
30 PCR Cycles
Plasmid containing
Ro52
Single Colony
<100'000 bacteria
Microtiter plate
Ref.
No PCR
10
20
30
Photon Intensity per PCR Cycle from a Single Colony
Photon Intensity (Photons/sec)
PCR Cycle